

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:40:09 ; Search time 26.7893 Seconds

(without alignments)
646.490 Million cell updates/sec

Title: US-10-699-035a-2

Perfect score: 913

Sequence: 1 RGDLMFLDSSASVSHYEPS.....FVDDHLHIVQLRGSILD 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	265.5	29.1	1857	2 S31212	collagen alpha 1(X)
2	265.5	29.1	1858	2 S74476	collagen alpha 1(X)
3	265.5	29.1	3124	2 A40020	collagen alpha 1(X)
4	254.5	27.9	1747	2 A45974	collagen alpha 1(X)
5	247.5	27.1	493	2 A33809	cartilage matrix P
6	239.5	26.2	500	2 S66222	cartilage matrix P
7	233.5	25.6	496	2 A37979	cartilage matrix P
8	227.5	24.9	2944	2 A54849	collagen alpha 1(V)
9	216.5	23.7	3137	2 A37797	collagen alpha 3(V)
10	214	23.4	741	2 T46488	hypothetical prote
11	207.5	22.7	3176	2 CGH3A	collagen alpha 3(V)
12	198.5	21.7	929	2 I51027	type XII collagen
13	198	21.7	272	2 A55348	integrin alpha-1 -
14	194	21.2	1151	2 A45226	integrin alpha-1 c
15	191.5	21.0	843	2 A40970	undulin 1 - human
16	184	20.2	1180	2 A35854	integrin alpha-1 c
17	179	19.6	550	2 T23760	hypothetical prote
18	175	19.2	724	2 A48569	antigen Em100 - Et
19	175	19.2	1286	2 A88396	protein M01E10.2 (
20	168	18.4	3051	2 S42373	hypothetical prote
21	162.5	17.8	1153	1 RWH1B	cell surface glyco
22	162.5	17.8	1170	2 I45914	integrin alpha 2 s
23	159.5	17.5	1153	2 S00551	leukocyte surface
24	158.5	17.4	1178	2 S44142	VLA-2 protein homo
25	158	17.3	712	2 A45638	immunodominant mic
26	153	16.8	1163	2 I56126	lymphocyte function
27	152.5	16.7	427	2 G00039	von Willebrand fac
28	151.5	16.6	1163	1 RWH1C	cell surface glyco
29	145.5	15.9	1181	2 A33998	integrin alpha-2 c

30	144.5	15.8	2813	1 VWHU	von Willebrand fac
31	140	15.3	414	2 P50323	von Willebrand fac
32	139.5	15.3	191	2 I47230	VLA-2 protein - pi
33	133.5	14.6	1179	2 A53213	integrin alpha-Ec
34	133	14.6	1170	2 S03308	cell surface glyco
35	113	12.4	640	2 T29784	hypothetical prote
36	111	12.2	13055	2 T16580	hypothetical prote
37	109.5	12.0	567	2 T28797	hypothetical prote
38	104	11.4	643	2 T19549	hypothetical prote
39	103.5	11.3	1028	1 CGH1A	collagen alpha 1(V
40	99.5	10.9	475	2 T36921	hypothetical prote
41	99.5	10.9	1019	1 A32856	collagen alpha 1(V
42	98.5	10.8	371	2 S32604	collagen alpha 2(V
43	96.5	10.6	1025	2 S34839	collagen alpha 1(V
44	96	10.5	918	2 S23377	collagen alpha 2(V
45	96	10.5	1022	2 S04111	collagen alpha 2(V

ALIGNMENTS

RESULT 1

S31212 collagen alpha 1(XIV) chain precursor, short form - chicken

C/Species: Gallus gallus (chicken)

C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 15-Sep-2003

C/Accession: S31212

R/Waelchli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.

Bur, J. Biochem 212, 483-490, 1993

A/Title: Complete primary structure of chicken collagen XIV.

A/Reference number: S31211; MUID:93185668; PMID:8444186

A/Accession: S31212

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: mRNA

A/Residues: 1-1857 <WAE>

A/Cross-references: UNIPARC:UPI000017123C; EMBL:X70792; NID:G288874; PIDN:CA50063.1; PI

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993

C/Genetics:

A/Gene: Col14A1

C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trinc

F.1-28/Domain: signal sequence #status predicted <SIG>

F.129-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <WAE>

F.129-110/Domain: fibronectin type III repeat homology <FN3A>

F.156-320/Domain: von Willebrand factor type A repeat homology <VWA1>

F.352-433/Domain: fibronectin type III repeat homology <FN3B>

F.442-525/Domain: fibronectin type III repeat homology <FN3C>

F.534-614/Domain: fibronectin type III repeat homology <FN3D>

F.623-707/Domain: fibronectin type III repeat homology <FN3E>

F.741-823/Domain: fibronectin type III repeat homology <FN3F>

F.832-914/Domain: fibronectin type III repeat homology <FN3G>

F.922-1009/Domain: fibronectin type III repeat homology <FN3H>

F.1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 29.1%; Score 265.5; DB 2; Length 1857;

Best Local Similarity 34.5%; Pred. No. 2.2e-15;

Matches 61; Conservative 33; Mismatches 82; Indels 1; Gaps 1;

QY	3	DLMFLDSSASVSHYEPSRVFVGLVAPLPLGALRASLVHVGSRPYTEFPFGSSS	62
DB	158	DIIVLVDSWSVSGRNFRLVRLLENLVSAFNVSEKTRVGLAQSGBPRTIMHINAYGT	217
QY	63	GRAADAVFASAKQMDTHTGILATVYAKEQLFAEASGARPGVKLVWTTDGGSSDPVPG	122
DB	218	KDAVDAVRNLPYKGNLTGLATLYIENSFKPEAGARPGVSKIGILITDQSDVDP	277
QY	123	PMQELKDGVTVFVISTGRGNFLELSAASAPAEKGLHFV-DVDDHLHIVQLRGS	178
DB	278	PAKNLRDAGIELFAIGVXNADINELKIASBDSFHYVNVADFNFMNSIVELTRTV	334

RESULT 2
S78476 collagen alpha 1(XIV) chain precursor, long form - chicken

C:Species: Gallus gallus (chicken)
C:Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 09-Jul-2004
C:Accession: S78476; S31211
R:Trueb, B.
Submitted to the EMBL Data Library, January 1993
A:Reference number: S78476
A:Accession: S78476
A:Molecule type: mRNA
A:Residues: 1-1888 <TRU>
A:Cross-references: UNIPROT:P32018; UNIPARC:UPI0000126031; EMBL:X70793; NID:G288872; PID:R:Maechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A:Title: Complete primary structure of chicken collagen XIV.
A:Reference number: S31211; MUID:93185668; PMID:8444186
A:Accession: S31211
A:Molecule type: mRNA
A:Status: preliminary
A:Residues: 1-416;1460-1811,1843-1888 <MAE>
A:Cross-references: UNIPARC:UPI00001773F4; UNIPARC:UPI00001773F5; EMBL:X70793
C:Genetics:
A:Gene: COL14A1
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <MAT>
F:29-110/Domain: fibronectin type III repeat homology <FN3A>
F:156-330/Domain: von Willebrand factor type A repeat homology <VWA1>
F:352-433/Domain: fibronectin type III repeat homology <FN3B>
F:442-525/Domain: fibronectin type III repeat homology <FN3D>
F:534-614/Domain: fibronectin type III repeat homology <FN3E>
F:623-707/Domain: fibronectin type III repeat homology <FN3F>
F:741-823/Domain: fibronectin type III repeat homology <FN3G>
F:833-914/Domain: fibronectin type III repeat homology <FN3H>
F:922-1009/Domain: fibronectin type III repeat homology <FN3I>
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 29.1%; Score 265.5; DB 2; Length 1888;
Best Local Similarity 34.5%; Pred. No. 2.3e-15;
Matches 61; Conservative 33; Mismatches 82; Indels 1; Gaps 1;

Qy 3 DLMFLDSSASVSHYEFGRVREFVQGLVAPLPGLGALRASIVHNGSRPYTEPPGQSS 62
Db 158 DIVILVDSWSIGRNFRLVLFLENLVSARFVSGSEKRVGLAQSGDPRFEMHINAYGT 217

Qy 63 GEAADAVRASQRMGDTHTGTALVYAKEQLEPAEASGARPGVKKLVWVTDGSSDPYGP 122
Db 218 KDAVLDVAVRNLPYKGNLTGLTLYILSNSFKPPAGARPGVSKIGITLTDGKSDVDYIP 277

Qy 123 PMQELKDIQVVFVSTGRGNLELSAASAPAEKHFFV-DVDDLHIIVQELRQSI 178
Db 278 PAKNLRDAGIELFALGVKNADINELKEIASSEPDSTHYVNAVDFMNSIVEGLTHTV 334

RESULT 3
A40020
collagen alpha 1(XII) chain precursor - chicken
N:Alternate names: fibrochimarin
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811
R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obari, J. Cell Biol. 115, 209-221, 1991
A:Title: The complete primary structure of type XII collagen shows a chimeric molecule within region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.
A:Reference number: A40020; MUID:92011862; PMID:1918137
A:Accession: A40020
A:Molecule type: mRNA
A:Residues: 1-3124 <YAM>
A:Cross-references: UNIPROT:P13944; UNIPARC:UPI000012602D; GB:D00824; NID:G222810; PIDN:A:Note: in the authors' translation residues 1216-1219 are shown after residue 1235 and, R:Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
J. Biol. Chem. 264, 19772-19778, 1989
A:Title: Type XII collagen. A large multidomain molecule with partial homology to type I
A:Reference number: A34485; MUID:90062079; PMID:2584192

A:Accession: A34485
A:Molecule type: mRNA
A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
A:Cross-references: UNIPARC:UPI0000171233; EMBL:J05137; NID:G211284; PIDN:AAA48635.1; PI
A:Accession: B34485
A:Molecule type: Protein
A:Residues: 2772-2792;2846-2873 <GOR2>
A:Cross-references: UNIPARC:UPI0000173C3F; UNIPARC:UPI0000173C40
R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A:Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA c
A:Reference number: A28037; MUID:87317590; PMID:3476925
A:Accession: A28037
A:Molecule type: mRNA
A:Residues: 2960-2976, 'F', 2978-3074, 'AG' <GOR3>
A:Cross-references: UNIPARC:UPI00001712F4; EMBL:M17375; NID:G211649; PIDN:AAA48718.1; PI
A:Note: this sequence has been revised in reference A34485
R:Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992
A:Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A:Reference number: S23814; MUID:92362621; PMID:1323460
A:Accession: S23814
A:Molecule type: Protein
A:Residues: 'X', 1333, 'Q', 1335-1347;1914-1928;2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517 <
A:Cross-references: UNIPARC:UPI0000173C41; UNIPARC:UPI0000173C42; UNIPARC:UPI0000173C43
R:Dublet, B.; van der Rest, M.
J. Biol. Chem. 262, 17724-17727, 1987
A:Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepain-
A:Reference number: S22254; MUID:88087065; PMID:33121603
A:Accession: S22254
A:Molecule type: protein
A:Residues: 2831-2833, 'T', 2834, 'R', 2836-2843;3002-3014 <DUB>
A:Cross-references: UNIPARC:UPI0000173C44; UNIPARC:UPI0000173C45
R:Trueb, J.; Trueb, B.
Biochim. Biophys. Acta 1171, 97-98, 1992
A:Title: The two splice variants of collagen XII share a common 5' end.
A:Reference number: S28811; MUID:93042014; PMID:1420368
A:Accession: S28811
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24,1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>
A:Cross-references: UNIPARC:UPI0000173C46; EMBL:X67327
C:Genetics:
A:introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disulfid
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
F:24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted
F:24-114/Domain: IITa #status predicted <IITa>
F:124-105/Domain: fibronectin type III repeat homology <FN3A>
F:133-301/Domain: von Willebrand factor type A repeat homology <VWA1>
F:332-425/Domain: IITB #status predicted <IITB>
F:333-414/Domain: fibronectin type III repeat homology <FN3B>
F:437-601/Domain: von Willebrand factor type A repeat homology <VWA2>
F:623-1178/Domain: IITC #status predicted <IITC>
F:630-711/Domain: fibronectin type III repeat homology <FN3C>
F:721-802/Domain: fibronectin type III repeat homology <FN3D>
F:812-895/Domain: fibronectin type III repeat homology <FN3E>
F:905-986/Domain: fibronectin type III repeat homology <FN3F>
F:995-1076/Domain: fibronectin type III repeat homology <FN3G>
F:1086-1169/Domain: fibronectin type III repeat homology <FN3H>
F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>
F:1384-1465/Domain: IITD #status predicted <IITD>
F:1384-1465/Domain: fibronectin type III repeat homology <FN3I>
F:1474-1557/Domain: fibronectin type III repeat homology <FN3J>
F:1566-1647/Domain: fibronectin type III repeat homology <FN3K>
F:1655-1738/Domain: fibronectin type III repeat homology <FN3L>
F:1756-1838/Domain: fibronectin type III repeat homology <FN3M>
F:1847-1928/Domain: fibronectin type III repeat homology <FN3N>
F:1937-2019/Domain: fibronectin type III repeat homology <FN3O>
F:2028-2110/Domain: fibronectin type III repeat homology <FN3P>
F:2119-2199/Domain: fibronectin type III repeat homology <FN3Q>
F:2207-2294/Domain: fibronectin type III repeat homology <FN3R>


```

Db      332  KKDIIKAAVKKKAAVNEKGMITGOALKLYLVDSFSLTANGARPEVPRKGVIFTDGRSQDIYTD 391
Oy      123  PMQELKDIGVTVFIVSTGRGNFLEISAAASAPAEKHLLFVDDV 165
       : ||||| : ||||| : ||||| : ||||| : |||||
Db      392  AAKRAKDIGFRMFPAVGVGNVAVEDELRELTASPEVAEH-YFYTAD 433

RESULT 6
S66552
cartilage matrix protein precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S66552
R:Ascoli, A.; Hauser, N.; Stender, D.; Paulsson, M.; Hitiipi, L.; Bosze, Z.
Eur. J. Biochem. 236, 970-977, 1996
A>Title: Cloning, sequencing and expression analysis of mouse cartilage matrix protein c
A:Reference number: S66522; MUID:96270751; PMID:8665920
A:Accession: S66522
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-500 <ASZ>
A:Cross-references: UNIPROT:P51942; UNIPARC:UPI0000029480; EMBL:U35035; NID:g1163176; PI
C:Genetic8:
A:Gene: CMP
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repeat
F.1-29/Domain: signal sequence #status predicted <SIG>
F.30-500/Product: cartilage matrix protein #status predicted <MAT>
F.43-210/Domain: von Willebrand factor type A repeat homology <WMA1>
F.231-266/Domain: EGF homology <EGF>
I:277-441/Domain: von Willebrand factor type A repeat homology <WMA2>

```

```

Query March 26.2% Score 239.5; DB 2; Length 500;
Best Local Similarity 34.8%; Pred. No. 1e-13; Mismatches 29; Gaps 1;
Matches 57; Conservative 29; Indels 3; Gaps 1;

Cy 3 DLMLDDSSASVSHYESFSRVREPVGOVLVAPLPLGTGALRASLVHVGSRPYTEPPFGHSS 62
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 45 DLVFPVDSRSRVRPEPEFKVKFVLSQVIESLDVGNATRVGLVNVVASTVKEPEPLRAHGS 104
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cy 63 GEAADDAVARASQNRGDPHTHGLALVYAKEQLFAEASGAR---PGVPKVLVWITGGSSDP 119
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 105 KASLIQARRRRIQPLSTGTMTGLALQFATKLSDBEGGRARSPDISKVIIVITDGRPPDS 164
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cy 120 VGPPMQLKDLGVTFVIFVISTGRGNFLLESAASAPAEKGLHFVD 163
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 165 VRDVERARASQIELFALIGLRVDPKATLRQLASEPQDEHDVYVE 208
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 7
A37979
cartilage matrix protein precursor - human
C:Species: Homo sapiens (man)
C:Date: 12-Jul-1991 #revision 12-Jul-1991 #text_change 09-Jul-2004
C:Accession: A37979; B37979
J:Jenkins, R.N.; Osborne-Lawrence, S.L.; Sindelar, A.K.; Eddy Jr., R.L.; Byere, M.G.; Sh
J: Biol. Chem. 265: 19624-19631, 1990
A:Title: Structure and chromosomal location of the human gene encoding cartilage matrix
A:Reference number: A37979; MUID:91060568; PMID:2246248
A:Accession: A37979
A:Molecule type: DNA
A:Residues: 1-496 <JEN>
A:Cross-references: UNIPROT:P21941; UNIPARC:UPI000004F1ED; GB:J05667
A:Accession: B37979
A:Molecule type: mRNA
A:Residues: 157-290, 'L', 292-496 <JEN>
A:Cross-references: UNIPARC:UPI000016A6E8; GB:M55683; GB:J05666; GB:J05667; NID:9180651
A:Gene: GDB:CRTM
A:References: GDB:127280; OMIM:115437
A:Map position: 1p35-1p35
A:Introns: 32/1; 147/3; 222/1; 264/1; 403/1; 454/1; 481/1
C:Complex: homotrimer
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe

```

C:Keywords: glycoprotein; homotrimer
F1-32/Domain: signal sequence #status predicted <SIG>
F123-496/Product: cartilage matrix protein #status predicted <MAT>
F139-206/Domain: von Willebrand factor type A repeat homology <VWA1>
F1227-262/Domain: EGF homology <EGF>
F1273-437/Domain: von Willebrand factor type A repeat homology <VWA2>
F176,344/Binding site: carbohydrate (Asn) (covalent) #status predicted
F1221-238,234-247,249-262/Disulfide bonds: #status predicted

[illegible]

RESULT 8

collagen alpha 1(VII) chain precursor - human
A:Accession: A54849

N/Alternate names: procollagen alpha 1(VII) chain
C/Species: Homo sapiens (man)
C/Date: 04-Nov-1994 #sequence revision 04-Nov-1994 #next change 09-Jul-2004
C/Accession: A58849; PH08844; S16316; I56328; A30266; I84686
R/Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.
J. Biol. Chem. 269, 20256-20262, 1994

A>Title: Cloning of human type VII collagen. Complete primary sequence of the alpha(VII)
A/Reference number: A54849; MUID:94327588; PMID:8051117

A/Accession: A54849

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-2944 <CHR>

A/Cross-references: UNIPROT:002388; UNIPARC:UPI000017A138; GB:L02870; NID:9987124; PIDN:7

R/Tanaka, T.; Takahashi, K.; Furukawa, F.; Imanura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992

A>Title: Molecular cloning and characterization of type VII collagen cDNA.
A/Reference number: PH0844; MUID:92231902; PMID:1567409

A/Accession: PH0844

A/Molecule type: mRNA

A/Residues: 'EPR',340-475, 'PALSTASHSLTCWRATRWGHCHNGSHWTRPACCEPCNPASHRAARAG',524-528 'C',5
A/Cross-references: UNIPARC:UPI000017A139; DDBJ:D11152; DDBJ:D13694; NID:9453658; PIDN:BA

A/Experimental source: Keratinocyte
A/Note: The authors translated the codon ACC for residues 394 and 397 as Tyr
R/Parente, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Matu
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991

A>Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A/Reference number: S16316; MUID:91334380; PMID:1871109

A/Accession: S16316

A/Molecule type: mRNA

A/Residues: 815-892, 'E',894-1439 <PAR>
A/Cross-references: UNIPARC:UPI000016A722; GB:M65158; GB:S49017; NID:gl80914; PIDN:AAA96

A/Experimental source: Keratinocyte
R/Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prissayanh, P.S.; Cook, M.E.; Wright, J.;
J. Invest. Dermatol. 99, 691-696, 1992

A>Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot
A/Reference number: I56328; MUID:93107742; PMID:1469284

A/Accession: I56328

A/Status: translated from GB/EMBL/DBDJ

A/Molecule type: mRNA

A/Residues: 'EPR',372-517, 'DV',520-540, 'W',542-1255 <RES>
A/Cross-references: UNIPARC:UPI000016B3AC; GB:S51236; NID:9262308; PIDN:AB24637.1; PID:6

J. Seltzer, J.L.; Ehsen, A.Z.; Bauer, E.A.; Morris, N.P.; Gnanville, R.W.; Burgeson, R.E.
J. Biol. Chem. 264, 3822-3826, 1989

A>Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase


```
Qy      63 GEAADAVRASAQRMGDTHTGLALVYAEQLEFAEASGAR--PGVPLVWVTGGSSDPV 120
      :      :      :      :      :      :      :      :      :      :
Db      98 NQDVLISHANNPYMGSGSKTGKGLLELLNHLTKAAGSGVQVITVLIDGQSQDDV 157
      :      :      :      :      :      :      :      :      :      :
Qy     121 GPPQELDLGVTVIVTGTGRNFIETLSAASAPAEKHL-HFVDDVDLHITVQELRGSI 178
      :      :      :      :      :      :      :      :      :      :
Db     158 ALPSSVLKSAHNMIAVGQDAVEGELKEIASRPDTHLFNDELFTALHGIADVLSV 216
      :      :      :      :      :      :      :      :      :      :

RESULT 10
T46488
hypothetical protein DKFZp434J065.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C/Accession: T46488
R/Dusterhoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A/Reference number: 223035
A/Accession: T46488
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-741 <AAA>
A/Cross-references: UNIPROT:O8NDE6; UNIPARC:UPI000016ACCB; EMBL:AL137638
C/Genetics:
A/Note: DKFZp434J065.1

Query Match      23.4%; Score 214; DB 2; Length 741;
Best Local Similarity 31.7%; Pred. No. 3, 1e-11;
Matches 58; Conservative 35; Mismatches 84; Indels 6; Gaps 4;

Qy      3 DLMLFLDSSASVSHYEFGRVREVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQSS 62
      :      :      :      :      :      :      :      :      :      :
Db     440 DLVEFIDSKSLGCENFVEVVKQFTGIDSLTISPKAAVGLQSTQVHTEFTLRNNS 499
      :      :      :      :      :      :      :      :      :      :
Qy      63 GEAADAVRASAQRMG-DTHGLALVYAEQLEFAEASGAR--GVPKVLWVTGGSSD 118
      :      :      :      :      :      :      :      :      :      :
Db     500 AKDMKAAV-AHMKYKSGSMGTGLAKHMERSEFTGEGARPLSTRPRAIVFTDGRAQD 558
      :      :      :      :      :      :      :      :      :      :
Qy     119 PVGPQMOELKDLGVTVIVTGTGRNFIETLSAASAPAEKHLHFV-DVDDLHITVQELRG 177
      :      :      :      :      :      :      :      :      :      :
Db     559 DVESEASAKAKNGITMYAVGVKALEEELQETASEPTNKLHYADPSTMDISKLLKG 618
      :      :      :      :      :      :      :      :      :      :
Qy     178 ILD 180
      :      :      :      :      :      :      :      :      :      :
Db     619 ICE 621

RESULT 11
CGH03A
collagen alpha 3(VI) chain precursor [validated] - human
N/Contains: collagen alpha 3(VI) chain, splice form A9/N10(-)
C/Species: Homo sapiens (man)
C/Date: 21-Nov-1993 #sequence_revision 12-Nov-1999 #text_change 16-Aug-2004
C/Accession: A51140; S13679; S24465; A57083; S28776; S00245; C31952; C29848; S26510; S48
R/Chu, M.L.
submitted to GenBank, May 1998
A/Reference number: A59140
A/Accession: A59140
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-3176 <CHU>
A/Cross-references: UNIPROT:P12111; UNIPARC:UPI0000126DAF; GB:X52022; NID:g3127925; PIDN
R/Chu, M.L.; Zhang, R.Z.; Pan, T.; Stokes, D.; Conway, D.; Kuo, H.J.; Glanville, R.; May
EMBO J. 9, 385-393, 1990
A/Title: Mosaic structure of globular domains in the human type VI collagen alpha-3 cha
itors.
A/Reference number: S13679; MUID:90151612; PMID:1689238
A/Accession: S13679
A/Molecule type: mRNA
A/Residues: 1-30,237-313, 'CWW', 318-322, 'AR', 326-1815, 'PD', 1818-1819, 'ID', 1822-3176 <CH5>
A/Cross-references: UNIPARC:UPI0000173C14; EMBL:X52022; NID:g3127925
A/Accession: S24465
```

```
A/Molecule type: protein
A/Residues: 574-585;965-973, 'X', 975-976;1306-1325;1361-1377;1381-1401;1473-1506, 'X', 1508
-1962, 'X', 1964-1965;2018-2037;2274-2410;2445-2459;2466-2469, 'X', 2471-2474;2504-2508, 'X',
A/Cross-references: UNIPARC:UPI0000173C15; UNIPARC:UPI0000173C16; UNIPARC:UPI0000173C17;
C/C: UNIPARC:UPI0000173C1D; UNIPARC:UPI0000173C1E; UNIPARC:UPI0000173C1F; UNIPARC:UPI0000
10000173C25; UNIPARC:UPI0000173C26; UNIPARC:UPI0000173C27
R/Zanussi, S.; Dollana, R.; Segat, D.; Bonaldo, P.; Colomatti, A.
J. Biol. Chem. 267, 24082-24089, 1992
A/Title: The human type VI collagen gene. mRNA and protein variants of the alpha3 chain
A/Reference number: S28776; MUID:93054780; PMID:1339440
A/Accession: A57083
A/Molecule type: DNA
A/Residues: 310-328 <ZAN>
A/Cross-references: UNIPARC:UPI0000173C28
A/Accession: S28776
A/Molecule type: mRNA
A/Residues: 32-126, 'AK', 129-136, 'L', 138-236 <ZAZ>
A/Cross-references: UNIPARC:UPI0000173C28; GB:SA9432; NID:g260296; PIDN:AA824261.1; PID:
R/Chu, M.L.; Mann, K.; Deutzmann, R.; Pribula-Conway, D.; Hsu-Chen, C.C.; Bernard, M.P.;
Eur. J. Biochem. 168, 309-317, 1987
A/Title: Characterization of three constituent chains of collagen type VI by peptide seq
A/Reference number: S00126; MUID:88029444; PMID:3665927
A/Accession: S00245
A/Molecule type: mRNA, protein
A/Residues: 2024-2046;2092-2156, 'R', 2203-2208, 'X', 2210-2211, 'X', 2213-2227;2228-2251;2314
A/Cross-references: UNIPARC:UPI000016A705; UNIPARC:UPI0000173C29; UNIPARC:UPI0000173C2A;
ID:g1335034
A/Note: the mRNA portion of the sequence corresponds to residues 2092-2157
R/Chu, M.L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutzmann, R.; Timpl, R.
J. Biol. Chem. 263, 18601-18606, 1988
A/Title: Amino acid sequence of the triple-helical domain of human collagen type VI.
A/Reference number: A31952; MUID:89066644; PMID:3198591
A/Accession: C31952
A/Molecule type: mRNA
A/Residues: 2038-2373 <CH4>
A/Cross-references: UNIPARC:UPI0000173C2E; GB:J04211; GB:M20778
A/Note: parts of this sequence were determined by protein sequencing
R/Well, D.; Mattei, M.G.; Passage, E.; Van Cong, N.; Pribula-Conway, D.; Mann, K.; Deutz
Am. J. Hum. Genet. 42, 435-445, 1988
A/Title: Cloning and chromosomal localization of human genes encoding the three chains of
A/Reference number: A29848; MUID:88161046; PMID:3348212
A/Accession: C29848
A/Molecule type: mRNA
A/Residues: 2092-2151 <WEI>
A/Cross-references: UNIPARC:UPI000016A720; GB:M27449; NID:g291919; PIDN:AA52057.1; PID:
A/Note: part of this sequence was determined by protein sequencing
R/Jander, R.; Rautenberg, J.; Glanville, R.W.
Eur. J. Biochem. 133, 39-46, 1983
A/Title: Further characterization of the three polypeptide chains of bovine and human sh
A/Reference number: S26506; MUID:83209648; PMID:6852033
A/Accession: S26510
A/Molecule type: protein
A/Residues: 'SALAGVAGVG' <JUN>
A/Cross-references: UNIPARC:UPI0000173C2F
A/Note: this sequence cannot be reliably placed and probably represents the residues from
R/Mayer, U.; Poeschel, E.; Nisch, R.; Specks, U.; Pan, T.C.; Chu, M.L.; Timpl, R.
Eur. J. Biochem. 225, 573-580, 1994
A/Title: Recombinant expression and properties of the Kunitz-type protease-inhibitor mod
A/Reference number: S48709; MUID:95045506; PMID:7525281
A/Accession: S48709
A/Molecule type: mRNA
A/Residues: 'MRAMIFFLCLAGALAA', 3102-3176 <MAY>
A/Cross-references: UNIPARC:UPI0000173C30
A/Note: engineered sequence to allow isolation of the Kunitz-type proteinase inhibitor h
R/Ronoux, B.; Merigoux, K.; Saladijan, P.; Norris, F.; Norris, K.; Bjorn, S.; Olsen, O.;
submitted to the Brookhaven Protein Data Bank, August 1994
A/Reference number: A52812; PDB:1IKU1
A/Contents: annotation; X-ray crystallography, 1.6 angstroms, residues 3106-3160
A/Note: engineered sequence expressed in Saccharomyces cerevisiae strain mc-663
C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
sines are 5-hydroxylated and subsequently O-glycosylated.
C/Comment: The fibronectin type III repeat homology domain may be released during proces
C/Genetics:
```

A:Gene: GDB:COL6A3
A:Cross-references: GDB:119066; OMIM:120250
A:Map position: 2q37.3-2q37.3
C:Complex: type VI collagen is a heterotrimer of one alpha 1(VI) chain (see PIR:CGH1A),
associations among trimer amino- and carboxyl-terminal domains (with disulfide bonds).
C:Function:
A:Description: structural component of extracellular tissue microfibrils associated with
C:Superfamily: collagen VI
C:Keywords: alternative splicing; blocked amino end; cell binding; coiled coil; extracel
F:1-25/Domain: signal sequence #status predicted <Sig>
F:26-3176/Product: collagen alpha 3(VI) chain #status predicted <AMT1>
F:26-3037/Domain: amino-terminal nonhelical #status predicted <AMT1>
F:26-30, 237-3176/Product: collagen alpha 3(VI) chain, splice form A9/N10(-) #status pred
F:337-203/Domain: von Willebrand factor type A repeat homology <VM01>
F:340-405/Domain: von Willebrand factor type A repeat homology <VM02>
F:443-608/Domain: von Willebrand factor type A repeat homology <VM03>
F:637-802/Domain: von Willebrand factor type A repeat homology <VM04>
F:835-999/Domain: von Willebrand factor type A repeat homology <VM05>
F:1027-1191/Domain: von Willebrand factor type A repeat homology <VM06>
F:1231-1394/Domain: von Willebrand factor type A repeat homology <VM07>
F:1433-1599/Domain: von Willebrand factor type A repeat homology <VM08>
F:1637-1802/Domain: von Willebrand factor type A repeat homology <VM09>
F:1836-2005/Domain: von Willebrand factor type A repeat homology <VM10>
F:2038-2373/Region: interrupted helical
F:2040-2042/Region: cell attachment (R-G-D) motif
F:2136-2138/Region: cell attachment (R-G-D) motif
F:2146-2150/Region: cell attachment (R-G-D) motif
F:2154-2156/Region: cell attachment (R-G-D) motif
F:2370-2372/Region: cell attachment (R-G-D) motif
F:2374-3176/Domain: carboxyl-terminal nonhelical #status predicted <CMH>
F:2400-2511/Domain: von Willebrand factor type A repeat homology <VM11>
F:2611-2800/Domain: von Willebrand factor type A repeat homology <VM12>
F:2865-2986/Region: von Willebrand factor type A repeat homology <VM13>
F:2987-3072/Domain: fibronectin type III repeat homology <FN3>
F:3111-3161/Domain: animal knittz-type proteinase inhibitor homology <BPI>
F:326/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:108,116,202,251,2079,2331,2558,2677,2861,3036/Binding site: carboxylate (Asn) (cova
F:2087/Disulfide bonds: Interchain #status predicted
F:2100,2206,2239,2316,2319/Modified site: 4-hydroxyproline (Pro) #status experimental
F:2103,2209,2212,2332,2337/Binding site: carboxylate (Lys) (covalent) #status experimen
F:2103,2209,2212,2332,2337/Modified site: 5-hydroxylysine (Lys) #status experimental
F:3111-3161,3120-3144,3136-3157/Disulfide bonds: #status predicted

Query Match 22.7%; Score 207.5; DB 2; Length 3176;
Best Local Similarity 31.4%; Pred. No. 6.5e-10;
Matches 55; Conservative 28; Mismatches 89; Indels 3; Gaps 2;

QY 3 DLMFLDSSASVSHYEFRRVFGVLVAPLPGTALRASLVHVSRRPYTEPPFGQHS 62
DB 39 DIFLVDSSWTIGEEHFOVREFLYDVVKSLLAVGENDFHFALVQFNGNPHTEFLNTYRT 98
QY 63 GEAADAVRASQRMG-DTHGTALVYAKEQLFAEASGAR--PGVPVLVWMTDGGSSDPV 120
DB 99 KQEVLSHINMSTYIGSTNQTGKLEIYMSHLLTKAASGASGCVPIVYVLLDGHSGKDL 158
QY 121 GPPMOELKDGLGVTVIVSTGRGNFLELSAASAPAEKHL-HFVDDHLIIYQEL 174
DB 159 ALPSAEKSGADVNFVFAIVEADDEGALKEIASPELWMHFNLENFTSLHDIQNL 213

RESULT 12
151027
type XII collagen alpha-1 chain - eastern newt (fragment)
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 03-Jul-2004
C:Accession: I51027
R:Wei, Y.; Yang, E.V.; Klatte, K.P.; Taseava, R.A.
Dev. Biol. 168, 503-513, 1995
A:Title: Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII collag
A:Reference number: I51027; MUID:95246925; PMID:7729585
A:Accession: I51027
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Residues: 1-929 <WE1>
A:Cross-references: UNIPROT:Q91145; UNIPARC:UPI000126D2E; EMBL:U19494; NID:G632647; PIR
F:155-236/Domain: fibronectin type III repeat homology <3FR>
F:631-795/Domain: von Willebrand factor type A repeat homology <VMA3>

Query Match 21.7%; Score 198.5; DB 2; Length 929;
Best Local Similarity 27.9%; Pred. No. 9.9e-10;
Matches 50; Conservative 37; Mismatches 91; Indels 1; Gaps 1;

QY 1 RDDMLFDSSASVSHYEFRRVFGVLVAPLPGTALRASLVHVSRRPYTEPPFGQ 60
DB 631 QNDIVLVDGSSISGRPPNFKIVNFI SRVVEFDIGSDRQJAVQYSGDPTEWQLNTH 690
QY 61 SSGEADAVRASQRMG-DTHGTALVYAKEQLFAEASGARPGVPVLVWMTDGGSSDPV 120
DB 691 KTKKSLMDAVANLPYKGGNTNTGSALEKFTLENNFPGVGMREKARKIAILTDGKSQDDI 750
QY 121 GPPMOELKDGLGVTVIVSTGRGNFLELSAASAPAEKHLHFV-DVDDHLIIYQELRGS 178
DB 751 VAPSRKYADEGIELYAVGIRKADENELKEIASDPDELVMYVAVDFSLTNIVNDLTENV 809

RESULT 13
A55348
Integrin alpha-1 - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004
C:Accession: A55348
R:Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.E.
J. Biol. Chem. 269, 22811-22816, 1994
A:Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1.
A:Reference number: A55348; MUID:94357930; PMID:7521332
A:Accession: A55348
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-272 <KER>
A:Cross-references: UNIPROT:O42094; UNIPARC:UPI00017A14E; GB:U10114
F:55-230/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 21.7%; Score 198; DB 2; Length 272;
Best Local Similarity 34.2%; Pred. No. 2.6e-10;
Matches 65; Conservative 26; Mismatches 75; Indels 24; Gaps 8;

QY 3 DLMFLDSSASVSHYEFRRVFGVLVAPLPGTALRASLVHVSRRPYTEPPFGQHS 62
DB 57 DIVFLDSSNST--YPMSSVTAFLNSLLRNMDDIGQOYGVGYVQGVYHFFLYNTYST 114
QY 63 GEAADAVRASQRMG-DTHGTALVYAKEQLFAEASGARPGVPVLVWMTDGGSSDPV 121
DB 115 TEVWDALRLIRQRCGTQMTALGIDTAREEAFTEAHGARCGVQKVMYVITDGEESHD--N 172
QY 122 PPMQEL-----KDLGVTVIV-----STGRGNF-----ELSAASAPAEKHLHFVDDV 166
DB 173 YRLQEVTDKCEDENIQRPAIALIGSYSRGNLSTKFEVSEIKSIASKPTEK--HFNVSDE 230
QY 167 --LHIIYQEL 174
DB 231 LALVTVIEAL 240

RESULT 14
A45226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45226
R:Briesewitz, R.; Bpstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceeleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:44:15 ; Search time 0.959464 Seconds
(without alignments)
1468.904 Million cell updates/sec

Title: US-10-699-035A-1

Perfect score: 1049

Sequence: 1 ggggacctgattctctcgc.....tgaagggtccattctcgcg 537

Scoring table: BLOSUM62
Xgapop 10 0 ' Xgapext 0.5
Fgapop 10 0 ' Fgapext 0.5
Delop 6.0 ' Delext 7.0

Searched: 97014 segs, 13122536 residues

Total number of hits satisfying chosen parameters: 194028

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ndp.model -DEV=xlp
-o=/abs/ABSGWEE spool/US10699035/runat_13022006_062453_25634/ndp_query.fasta_1
-DB=Published Applications_AA_New -QPMT=fastan -SUFFIX=rapbn -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abs92p
-USER=US10699035 @CGN 1 1 10 @runat_13022006_062453_25634 -NCPU=6 -ICPU=3
-NO MAP -NEC SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications_AA_New:*

1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	902	86.0	445	6	US-10-453-372-2
2	261.5	24.9	3063	7	US-11-186-284-26
3	257.5	23.5	517	7	US-11-169-041-160
4	246.5	23.5	915	6	US-10-131-826A-294
5	246.5	23.5	956	7	US-11-113-424-39
6	197	18.8	214	7	US-11-192-449-6
7	197	18.8	214	7	US-11-192-449-9
8	188	17.9	214	7	US-11-192-449-5
9	183	17.4	678	6	US-10-063-703-34

10	183	17.4	678	7	US-11-102-240-34	Sequence 34, Appli
11	166.5	15.9	1152	7	US-11-080-026-4	Sequence 4, Appli
12	162	15.4	709	6	US-10-453-372-186	Sequence 186, App
13	161	15.3	709	6	US-10-453-372-180	Sequence 180, App
14	161	15.3	3568	6	US-10-453-372-194	Sequence 194, App
15	161	15.3	3570	6	US-10-453-372-178	Sequence 178, App
16	161	15.3	3570	6	US-10-453-372-196	Sequence 196, App
17	161	15.3	3570	6	US-10-453-372-198	Sequence 198, App
18	161	15.3	3570	6	US-10-453-372-200	Sequence 200, App
19	161	15.3	3570	6	US-10-453-372-202	Sequence 202, App
20	161	15.3	3570	6	US-10-453-372-204	Sequence 204, App
21	161	15.3	3570	6	US-10-453-372-206	Sequence 206, App
22	160	15.3	709	6	US-10-453-372-182	Sequence 182, App
23	155	14.8	709	6	US-10-453-372-184	Sequence 184, App
24	150.5	14.3	1167	6	US-10-601-368-18	Sequence 18, Appl
25	150	14.3	1141	6	US-10-601-368-24	Sequence 24, Appl
26	150	14.3	1166	6	US-10-601-368-22	Sequence 22, Appl
27	150	14.3	1188	6	US-10-601-368-21	Sequence 21, Appl
28	149.5	14.3	1182	6	US-10-601-368-25	Sequence 25, Appl
29	145.5	13.9	1147	6	US-10-453-372-4	Sequence 4, Appli
30	145	13.8	1141	6	US-10-601-368-6	Sequence 6, Appli
31	145	13.8	1166	6	US-10-601-368-4	Sequence 4, Appli
32	145	13.8	1188	6	US-10-601-368-3	Sequence 3, Appli
33	145	13.8	1188	7	US-11-000-463-338	Sequence 338, App
34	145	13.8	1188	7	US-11-000-463-810	Sequence 810, App
35	144.5	13.8	1862	6	US-10-601-368-7	Sequence 7, Appli
36	144.5	13.8	2764	6	US-10-995-561-691	Sequence 691, App
37	144.5	13.8	2813	6	US-10-995-561-688	Sequence 688, App
38	144.5	13.8	2919	6	US-10-821-234-1133	Sequence 1133, Ap
39	134.5	12.8	1179	7	US-11-097-125-1	Sequence 1, Appli
40	134.5	12.8	1196	6	US-10-995-561-921	Sequence 921, App
41	134	12.8	184	6	US-10-665-658-8	Sequence 8, Appli
42	133	12.7	184	6	US-10-665-658-7	Sequence 7, Appli
43	133	12.7	1170	7	US-11-080-026-2	Sequence 2, Appli
44	133	12.7	1170	7	US-11-107-028-4	Sequence 4, Appli
45	128	12.2	1167	7	US-11-097-125-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-453-372-2
; Sequence 2, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789290
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuroSeqList version 0.1

SEQ ID NO 2
LENGTH: 445
TYPE: PRT
ORGANISM: Homo sapiens
US-10-453-372-2

Alignment Scores:
Pred. No.: 1.63e-62 Length: 445
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 6 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-453-372-2 (1-445)

QY 1 GGGGACCGATGTTCTCTGTCGAGCAGTCCAGCCAGTCTCTCACTAGAGTTCTCCGG 60
DB 33 G1YsPLeuMePheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCGGAGATTGTTGGGCGAGCTGTGCTCCACTGCCCCCTGGGACCGGGCCCTGCGT 120
DB 53 ValArgGluPheValG1YGlneValAlaProLeuProLeuG1YThrG1YAlaLeuArg 72
QY 121 GCCACTGTGTGCACGTGTGGCGAGTCCGCCATACCCGAGTTCCCTTGGCCAGCAGC 180
DB 73 AlaSerLeuValHisValG1YSerArgProTyrThrGluPheProPheG1YHisSer 92
QY 181 TCGGGTAGGCTGCCCGAGATGCGGTGGCTTCTGCGCCAGGATGGGTAGACCCAG 240
DB 93 SerG1YValAlaAlaGlnsPalAvalAlaArgAlaSerAlaArgMetG1YAspThrHis 112
QY 241 ACTGGCTGTGGCTGTATGTCGAAGAACAGCTGTTTGTGAAGCATCAAGTCCCG 300
DB 113 ThrG1YLeuAlaLeuValTyrAlaYsg1YLeuPheAlaGluAlaSerG1YAlaArg 132
QY 301 CCAGGGGTCGCCAAAGTGTGTGTGGTGCACATGGCGGCTCCAGGACCCCTGTGGC 360
DB 133 ProG1YValProLYsValLeuValTyrValThrAspG1YG1YSerSerAspProValG1Y 152
QY 361 CCCCCAGCAGGACCTCAAGGACCTGGGGCGTCACCGGTTTATTGTGACGACCGGCGA 420
DB 153 ProPheG1YLeuLeuLeuAspLeuG1YValThrValPheLeuValSerThrG1YArg 172
QY 421 GGCAATTCTGTGACCTGTCAAGCCGCTGCTCAGCCCTCCGAGAACGCTGCATT 480
DB 173 G1YAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLeuHisLeuHisPhe 192
QY 481 GTGACGTGATGACCTGCACATCATTTGCCAAGAGCTGAGGGGCTCCATTCTC 534
DB 193 ValAspValAspAspLeuHisLeuValG1YLeuArgG1YSerLeuLeu 210

RESULT 2
US-11-186-284-26
Sequence 26, Application US/11186284
Publication No. US20050266493A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
APPLICANT: BURGART, Lawrence J.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: MPW01-029P2RNM
CURRENT APPLICATION NUMBER: US/11/186,284
CURRENT FILING DATE: 2005-07-21
PRIOR APPLICATION NUMBER: US/10/301,822
PRIOR FILING DATE: 2002-11-21

PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 3063
TYPE: PRT
ORGANISM: Homo sapiens
US-11-186-284-26

Alignment Scores:
Pred. No.: 2.2e-13 Length: 3063
Score: 261.50 Matches: 64
Percent Similarity: 52.5% Conservative: 29
Best Local Similarity: 36.2% Mismatches: 83
Query Match: 24.9% Indels: 1
DB: 7 Gaps: 1

US-10-699-035A-1 (1-537) x US-11-186-284-26 (1-3063)

QY 4 GACCTGATGTTCTGTGTCGAGCAGTCCAGCCAGCTCTCTCACTAGAGTTCTCCGGATT 63
DB 440 AspLeuValAlaPheLeuValAspG1YSerTyrSerLeuG1YLeuAlaAsnPheValLeuVal 459
QY 64 CGGAGATTGTTGGGCGAGCTGTGCTCCACTGCCCCCTGGGACCGGGCCCTGCGTCC 123
DB 460 ArgAlaPheLeuGluValLeuValLysSerPheGluLeuSerProAsnArgValGlnLe 479
QY 124 AGTGTGTGCACGTGTGGCGAGTCCGCATACCCGAGTTCCCTTGGCCAGCAGCAGCTCG 183
DB 480 SerLeuValGlnTyrSerArgAspProHisThrGluPheThrLeuLYsPheThrLYs 499
QY 184 GGTGAGCTGCCCAAGATGCGGTGCGTCTTCTGCCAGCCATGGGTGAACCCAGCACT 243
DB 500 ValGlnsPheLeuLeuG1YLeuAlaAlaAsnThrPheProTyrArgG1YG1YSerThrAsnThr 519
QY 244 GACCTGAGCTGTGTTATGTCGAAGAACAGCTGTTTGTGAAGCATCAGTCCCGGCGCA 303
DB 520 G1YysAlaMetThrTyrValArgGluLYsLeuPheValProSerLYsG1YSerArgSer 539
QY 304 GGGGTGCCCAAGTGTGTGTGGTGCACAGATGGCGGCTCCAGACGACCTGTGGGCCCC 363
DB 540 AsnValProLYsValMetLeuLeuLeuThrAspG1YLysSerSerAspAlaPheArgAsp 559
QY 364 CCATGACGAGCTCAAGGACCTGGGCGTCAACCGTTTCATTGTTCAGCAGCCGCGAGGC 423
DB 560 ProAlaLeuLYsLeuArgAsnSerAspValGluLeuPheAlaValG1YValLYsAspAla 579
QY 424 AACTTCCTGAGCTGTGTCAGCCGCTGCTCAGCCCTCCGAGAACGCTCCGACTTTGTG 483
DB 580 ValArgSerGluLeuGluAlaAlaAlaSerProAlaGluThrHisValPheThrVal 599
QY 484 ---GACGTGATGACCTGCACATCATTTGCCAAGAGCTGAGGGGCTCCATT 531
DB 600 GluAspPheAspAlaPheGlnArgL1eSerPheGluLeuThrGlnSerL1e 616

RESULT 3
US-11-169-041-160
Sequence 160, Application US/11169041
Publication No. US20060019284A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
FILE REFERENCE: 10001 NP
CURRENT APPLICATION NUMBER: US/11/169,041
CURRENT FILING DATE: 2005-06-28

```

; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 160
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-169-041-160

Alignment Scores:
Pred. No.: 4,94e-13 Length: 517
Score: 257.50 Matches: 64
Percent Similarity: 52.0% Conservative: 27
Best Local Similarity: 36.6% Mismatches: 83
Query Match: 24.5% Indels: 1
DB: Gaps: 1

US-10-699-035a-1 (1-537) x US-11-169-041-160 (1-517)
QY 10 ATGTCCTGCTGACAGCTCAGCCAGCGTCTCTCACTACGAGTTCTCCGGGTTCCGAG 69
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 143 MetPheLeuValAspIlySerIlySerIlyIleGlyIleAlaAsnPhelValIlyValAla 162
QY 70 TTTGTCGGGCACTGTGCTCCTCCTGCGCCCTGGGACCGGGGCTGCGTGCAGTCTG 129
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 PheLeuGluValIleuValIlySerPheGluIleSerProAsnArgValGlnIleSerLeu 182
QY 130 GTGCACTGTGGGACGCGCCATACACCGAGTTCCCTTGGCCGACAGACAGCTCGGTGAG 189
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 ValGlnIlySerArgAsnProIlySerGlnPheThrLeuIlySerPheThrIlyValGlu 202
QY 190 GCTGCCAGGATGCGGTGCTGCTTGGCCAGCGGATGGGTGACACCCACACTGCGCTG 249
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 203 AspIleIleGluAlaIleAsnThrPheProIlyArgGlyGlySerThrAsnThrIlyLys 222
QY 250 GCGCTGCTATGCCAGGAACAGCTGTTCTGTAACATCAGTGCCTCCGCGCCAGGGGTG 309
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 223 AlaMetThrIlyValArgGluIlyIlePheValProSerIlySerIlySerIlyVal 242
QY 310 CCCAAGTCTGCTGTGCGGTGACAGATGGCGGCTCCAGGACCCCTTGCGGCCCGCCAG 369
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 ProIlyValMetIleuIleThrAspGlyIlySerSerAspAlaPheArgAspProAla 262
QY 370 CAGGACTCAAGGACCTGGGCGCTGACCGGTTTCATGTGACAGCCGCGGACCACTTC 429
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 263 IleIlyLeuArgAsnSerAspValGluIlePheAlaValGlyValIlyAspAlaValAsp 282
QY 430 CTGAGCTGTCAAGCCGCTGCTGAGCCCTGCGGAGAACACCTGCACTTTGTG--GAC 486
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 283 SerGluLeuGluAlaIleAlaSerProProIlaGluThrIleValPheThrValGluAsp 302
QY 487 GTGAGTACCTGCATCATTTGTCCAAAGCTGAGGGGCTCCATT 531
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 PheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIle 317

RESULT 4
US-10-131-826a-294
; Sequence 294, Application US/10131826a
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
```

```

; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 294
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826a-294

Alignment Scores:
Pred. No.: 3,34e-12 Length: 915
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
DB: Gaps: 2

US-10-699-035a-1 (1-537) x US-10-131-826a-294 (1-915)
QY 4 GACCTGATGTTCTGCTGACAGCTCAGCCAGCGTCTCTCACTACGAGTTCTCCGGGTT 63
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 AspLeuValPheIleIleAsnSerSerArgSerValAsnThrIleAspIlyAlaIlyVal 76
QY 64 CCGGAGTTGTGGGGGACGTGTGCTGCCTGCGCCCTGGGACCGGGGCTGCGTGC 123
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 77 IysGluPheIleValAspIleLeuGlnPheLeuAspIleGlyProAspValThrArgVal 96
QY 124 AGTCTGTCAGCTGTCAGTGCAGCTGACGAGTTCCTTCCCTTGGCCAGGACAGCTG 183
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 97 GlyLeuLeuGlnIlyGlySerThrValIlyAsnGlnPheSerIleuIlyThrPheIlyAsp 116
QY 184 GGTGAGCTGCCAGAGATGCGGTGCTGCTTGGCCAGGACGAGTGTGACACCCACT 243
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 117 LysSerGluValIgluArgAlaValIlyAspArgMetArgHleuSerThrGlyThrMetThr 136
QY 244 GCGCTGCGCTGCTATGCCAAGGAACAGCTGTTCTGTAAGCATCAGTGCCTCCGCGCA 303
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 137 GlyLeuAlaIleGlnIlyAlaLeuAsnIleAlaPheSerGlnAlaGluIlyAlaArgPro 156
QY 304 -----GGGGTCCCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 354
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 157 LeuArgGluAsnValProArgValIleMetIleValIlnAspGlyArgProGlnAspSer 176
QY 355 GTGGGCCCCCATGACGAGAGCTCAAGGACCTGGGGGTGACCGGTTCATTGTTCAGCACC 414
```

Db 177 ValaIaGuValaIaAlaIaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
Qy 415 GCGCGAGCACTTCCTGAGAGCTGTCAGCCGCTGCTCAGCCCGCTGAGAGCACTG 474
Db 197 GlyInValaAspPheAsnThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216
Qy 475 -----CACTTTGTGAGCGTGAGTACCTGCATCATTTGTCACAGAG 516
Db 217 PheIeuValaIaAsnPheSerGlnIleGluThrLeuThrSerValaPheGlnLys 234

RESULT 5

US-11-113-424-39
; Sequence 39, Application US/11113424
; Publication No. US20050260713a1
; GENERAL INFORMATION:
; APPLICANT: Gangoli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 39
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-39

Alignment Scores:
Pred. No.: 3,34e-12 Length: 956
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
Gaps: 2
DB: 7

US-10-699-035A-1 (1-537) x US-11-113-424-39 (1-956)

Qy 4 GACCTGATGTTCTCTGCTGAGCAGCTCAGCGTCTCTACTAGAGTTTCCGGGTT 63
Db 57 AsplleuValaPheIleIleAspSerSerIaArgSerValaAsnThrIleAspTyrAlaLysVal 76
Qy 64 CGGAGTTTGGGGCAGCTGTCGCTCCACTGCCCCCTGGGACCGGGGCGCTGCGTCC 123
Db 77 LysGluPheIleIleValaAspIleLeuGlnPheLeuAspIleGlyProAspValaThrArgVal 96
Qy 124 AGTCGTGTCAGCTGCGGAGCTGCGCCATACACCGAGTTTCCCTTGGCGGACAGACTCG 183
Db 97 GlyLeuLeuGlnTyrGlySerThrValaLysAsnGluPheSerLeuLysThrPheLysArg 116
Qy 184 GGTGAGGCTGCCAGATGCGGTGCTTCTGCCACAGGATGGGTGACACCACT 243
Db 117 LysSerGluValaGluArgAlaValaLysArgMetArgHisLeuSerThrGlyThrMetThr 136

Qy 244 GCGCTGGCGCTGTCGTATGCCAAGAACGCTGTTTGTGTAGAGCATAGTCCCGGCCA 303
Db 137 GlyIeuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluAlaGluValaArgPro 156
Qy 304 -----GGGTCGCCAAGTGTGTCGTGGTGGTGCAGATGCGGCTCCAGGACCT 354
Db 157 LeuArgGluAsnValaProArgValaIleMetIleValaThrAspGlyArgProGlnAspSer 176
Qy 355 GTGGGCCCCCGGACGAGCAAGCTCAAGACCTGCGGCGCTCAGCGTTTCAATTTGACGACC 414
Db 177 ValaIaGuValaIaAlaIaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
Qy 415 GCGCGAGCACTTCCTGAGAGCTGTCAAGCCGCTGCTCAGCCCTGCCGAGAACGACCTG 474
Db 197 GlyInValaAspPheAsnThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216
Qy 475 -----CACTTTGTGAGCGTGAGTACCTGCATCATTTGTCACAGAG 516
Db 217 PheIeuValaIaAsnPheSerGlnIleGluThrLeuThrSerValaPheGlnLys 234

RESULT 6

US-11-192-449-6
; Sequence 6, Application US/11192449
; Publication No. US2005028181B1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Koteliensky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USCN3
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-192-449-6

Alignment Scores:
Pred. No.: 2.3e-08 Length: 214
Score: 197.00 Matches: 65
Percent Similarity: 47.9% Conservative: 28
Best Local Similarity: 33.5% Mismatches: 81
Query Match: 18.8% Indels: 20
Gaps: 7
DB: 7

US-10-699-035A-1 (1-537) x US-11-192-449-6 (1-214)

Qy 4 GACCTGATGTTCTCTGCTGAGCAGCTCAGCGTCTCTACTAGAGTTTCCGGGTT 63
Db 22 AsplleuValaIleValaAspGlySerAsnSerIle-----TyrProTyrAspSerVal 39
Qy 64 CGGAGTTTGGGGCAGCTGTCGCTCCACTGCCCCCTGGGACCGGGGCGCTGCGTCC 123
Db 40 ThrAlaPheLeuAsnAspLeuLysArgMetAspIleGlyProLysGlnThrGlnVal 59
Qy 124 AGTCGTGTCAGCTGCGGAGCTGCGCCATACACCGAGTTTCCCTTGGCGGACAGACTCG 183
Db 60 GlyIleValaGlnTyrGlyGluAsnValaThrHisGluPheAsnLeuAsnLysTyrSerSer 79
Qy 184 GGTGAGGCTGCCAGATGCGGTGCTTCTGCCACAGGCGCATGGGT--GACACCCAC 240


```

OY      4 GACCTGATGTTCTCTGTGGACAGCTGACGCGGTCTTCCATCCAGATGTTCCCGGTT 63
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      22 AsplIleValIleValIleuAspIleGlySerIleAsnSerIle-----TyrProIrrGluSerVal 39
OY      64 CGGAGATTGTGGGGCAGCTGTGTGAGCTCCATGCTCCCTGGGACCCGGGCGCTGGCC 123
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      40 IleAlaPheLeuAsnAspLeuLeuIlyAsrGmeAspIleGlyProIlysgInThrGluVal 59
OY      124 AGTCGTGGACAGTGGGGCAGTGGGCCCATACACGAGATTCCCTTGGGCCAGCAGCTCG 183
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      60 GlyIleValGlnTyrGlyGluAsnValThrHisGlnIleAsnLeuAsnIlyrIySerSer 79
OY      184 GGTGAGGCTGCCAGAGATCGCGGTGCTGTCTTCCACGCGCATGGGT--GACACCCAC 240
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      80 ThrGluGluValIleuValAlaIleAlaIlySylIleGlyArgGlnGlyGluGlnThrMet 99
OY      241 ACTGGCCTGGCGCGCTGTCTATGTCCCAAGAACAGCTGTTTGTCTAAGCATGAGTGCCTGG 300
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      100 ThrAlaLeuGlyIleAspThrAlaAsrIlysgIuAlaPheThrIluAlaArgIlyAlaArg 119
OY      301 CCAGGAGGTGCCCAAAGTGTGTGGTGGTGGAGACAGATGCGCGCTCCAGCGAC-----CCT 354
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      120 ArgGlyValIlybIyValMetValIleValThrAspGlyGluSerHisAspAsnTyrArg 139
OY      355 GTGGGCCCCCATGACGAGAGCTCAAGACCTGGGCGTCAACCGTTCATTGTCAGACACC 414
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      140 LeuIysgIuValIleGlnAspCysGluAspGluAsnIleGlnIAsrPheSerIleAlaIle 159
OY      415 -----GGCGGAGCAACTTCCTCG-----GAGCTGTACGCC 444
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      160 LeuGlyHisTyrAsnIAsrGlyAsnLeuSerThrGlyIlePheValGluGluIleIySer 179
OY      445 GCTGCCTACAGCCCTGCGCAGAAACACCTGCACTTTGTGGACGTGATGAC----- 495
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      180 IleAlaSerGluProThrGluIyS-----HisPheAsnValSerAspGluLeuAla 197
OY      496 CTGCACATCATGTCTCCAAAGCTGAGAGGGCTCCATTCTCGG 537
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      198 LeuValThrIleValIyAlaLeuGlyGluArgIlePheAla 211

RESULT 9
US-10-063-703-34
; Sequence 34, Application US/10063703
; Publication No. US20060008901A1
; GENERAL INFORMATION:
;   APPLICANT: Eaton, Dan L.
;   APPLICANT: Filvaroff, Ellen
;   APPLICANT: Geritsen, Mary E.
;   APPLICANT: Goddard, Audrey
;   APPLICANT: Godowski, Paul J.
;   APPLICANT: Grimaldi, Christopher J.
;   APPLICANT: Gurney, Austin L.
;   APPLICANT: Watanabe, Colin K.
;   APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,703
; PRIOR FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 34
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-703-34

```

Alignment Scores:	
Pred. No.:	2,556-07
Score:	183.00
Percent Similarity:	45.1%
Best Local Similarity:	30.6%
	Length: 67
	Matches: 53
	Conservative: 25
	Mismatches: 91

Query Match:	17.4%	Indels:	4
DB:	6	Gaps:	2
US-10-699-035A-1 (1-537) × US-10-063-703-34 (1-678)			

QY 4 GACCGATGTTCTTCGTGGACAGCTCAGCCAGGCTCTTCACATCGAGATTTCTCCGGGTT 63
 Db 495 AspllleGlyPheValIleAspGlySerSerValGlyThrGlyAsnPheArgThrVal 514
 QY 64 CGGAGATTGTGGGCGAGCTGTGGCTCCATCCGCCCTGGAGACCGGGGCCCTGGTGC 123
 Db 515 LeuGlnPheValIThrAsnLeuThrLysGluPheGluLysLeuSerAProThrArgIle 534
 QY 124 AGTCGGTCGACGTGGGCGAGTCGGCCATACACCGAGTTCCTCCGGCAGACAGCTCG 183
 Db 535 GlyAlaValGlnTyrThrYrGluGlnArgLeuGluPheGlyPheAspLysTyrSer 554
 QY 184 GGTGAGGCTGCCCGAGATGCGGTGCGTCTTGCACCAGGCGCATGGGTGACACCCCACT 243
 Db 555 LysProAspIleLeuAsnAlaIleLysArgValGlyTyrTrpSerGlyGlyThrSerThr 574
 QY 244 GGCCCTGGCGCTGTGTCATGCGACAGAGCTGTTTGCTGTAAGCATCAGTGGCCCGGCA 303
 Db 575 GlyAlaAlaIleAsnPheAlaLeuGluGlnPheLys-----SerLysPro 591
 QY 304 GGGGTGCCCAAGTGTGTGTGGTGAGTACAGATGCGCGCTCCAGACGACCTGTGGGCC 363
 Db 592 AsnLysArgLysLeuMetIleLeuIleThrAspGlyArgSerTyrAspAspValArgIle 611
 QY 364 CCCATGACAGAGCTCAAGAGCTGGGCGGTACACCGTGTTCATGTTCAGACCGCGGAGGC 423
 Db 612 ProAlaMetAlaAlaHisLeuLysGlyValIleIleThrYrAlaIleGlyValAlaTrpAla 631
 QY 424 AACTTCCTGGAGCTGTCAAGCGGCTGCTCAGCCCTTGCACGAGAGCACTTCACATTTTG 483
 Db 632 AlaGlnGluGluLeuGluValIleAlaThrHisProAlaArgAspHisSerPhePheVal 651
 QY 484 GAC---GTGATGACCTGCACATCATTTGTCCMAAGCTG 519
 Db 652 AspGluPheAspAsnLeuHisGlnTyrValProArgIle 664

```

RESULT 10
US-11-102-240-34
; Sequence 34, Application US/11102240
; Publication No. US20050260647A1
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Wood, William L.
; TITLE OF INVENTION: ANTIBODIES TO POLYPEPTIDES ENCODED BY A NUCLEIC ACID UNDEREXPRESS
; FILE REFERENCE: P3230R1C106C
; CURRENT APPLICATION NUMBER: US/11/102,240
; PRIOR FILING DATE: 2005-04-08
; PRIOR APPLICATION NUMBER: 10/063662
; PRIOR FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: 10/006867
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 60/170262
; PRIOR FILING DATE: 199-12-09
; NUMBER OF SEQ ID NOS: 170

```

```

; ORGANISM: Homo Sapien
US-11-102-240-34

Alignment Scores:
Pred. No.:      2.55e-07      Length:      678

```

Score:	45.1%	Matches:	53
Best Similarity:	18.3%	Conservative:	25
Best Local Similarity:	30.6%	Mismatches:	91
Query Match:	17.4%	Indels:	4
DB:	7	Gaps:	2

US-10-699-035A-1 (1-537) X US-11-102-240-34 (1-678)

RESULT 11

```

US-11-080-026--4
Sequence 4, Application US/11080026
Publication No. US20050260192A1
GENERAL INFORMATION:
APPLICANT: Springer, Timothy A.
APPLICANT: Shimaoka, Motomu
APPLICANT: Lu, Chafien
TITLE OF INVENTION: MODIFIED POLYPEPTIDES STABILIZED IN A
FILE REFERENCE: CFBF-P02-021
CURRENT APPLICATION NUMBER: US/11/080,026
PRIOR FILING DATE: 2005-03-15
PRIOR APPLICATION NUMBER: 09/945,265
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/229,700
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1152
TYPE: PRT
ORGANISM: Homo sapiens
US-11-080-026--4

```

Best Local Similarity:	28.3%	Mismatches:	92
Query Match:	15.9%	Indels:	11

US-10-699-035A-1 (1-537) X US-11-080-026-4 (1-1152)

RESULT 12
HS-10-453

Sequence 186, Application US/10453372
Publication No. US2006003323A1
GENERAL INFORMATION:
APPLICANT: Alsbroock, et al.
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
FILE REFERENCE: 21402-589 A
CURRENT APPLICATION NUMBER: US/10/453,372
CURRENT FILING DATE: 2003-06-03
PRIOR APPLICATION NUMBER: 09/789330
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: 60/185967
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 09/823187
PRIOR FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195792
PRIOR FILING DATE: 2000-03-10
PRIOR APPLICATION NUMBER: 09/839446
PRIOR FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: 60/199476
PRIOR FILING DATE: 2000-03-25
PRIOR APPLICATION NUMBER: 09/863776
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 60/208253
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 09/939398

```

PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: 60/227800
PRIOR FILING DATE: 2000-08-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1609
SOFTWARE: CurSeqList version 0.1
SEQ ID NO 186
LENGTH: 709
TYPE: PRT
ORGANISM: Homo sapiens
US-10-453-372-186

Alignment Scores:
Pred. No.: 1,04e-05 Length: 709
Score: 162.00 Matches: 53
Percent Similarity: 47.1% Conservative: 27
Best Local Similarity: 31.2% Mismatches: 74
Query Match: 15.4% Indels: 16
DB: Gaps: 5

US-10-699-035A-1 (1-537) x US-10-453-372-186 (1-709)
OY      4   GACCTGATGTTCTCCGTGGAGACGCTCAGCCAGCGTCTTCACTAAGAATTCTCCGGGATT    63
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       68   GluIeuValPheIeuValAspAspSerSerSerValIGluValAlaPheArgSerGlu    87
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
OY      64   CGGAGTTTGAGGGCAGCTGTGGCTGCACCTGCCCTCGGACCGGGGCCCTGGTCCC    123
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       88   LeuMetPheValArgLyLeuIeuSerAspPheProValValProThrAlaThrArgVal    107
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
OY     124   AGCTGCTGTCAGCACTGGGCGCATCCGACATACCAGTAGTCCCTTCCTGGCCAGACACTCG    183
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     108   AlaIleValThrPheSerSerLyAsnTrVal--ValProArgAlaAspTryIleSer    126
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
OY     184   GGTGAGCTGCCAG-----GATGCGGTGCGGCTTGCC    219
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     127   ThrArgAlaArgAlaArgGlnHisLysCysAlaLeuIeuLeuGlnGlnIleProAlaIleSer    146
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
OY     220   CAGCGCATGGGTGAACCCACACACTGGCGCTGTCTATGCCAAAGAACACTGTTT    279
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     147   TyrArgGlyGlyGlyThrTyrrThrylserGlyAlaPheGlnGlnAlaGlnIleLeuIeu    166
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
OY     280   GCTGAAGCATCAGGTGCCGCCGACGGGTGCCCAAAGTCTGGTGGGTACAGATGCC    339
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     167   -----HisAlaArgGlnAsnSerAlaIysValValPheLeuIleThrAspGly    182
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
OY     340   GGCTTCAGC-----GACCTGTGGGGCCCCCACATGACGAGGACCAAGCACTGGGCGTC    393
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     183   TyrSerAlaGlyGlyAspPro--ArgProIleAlaAlaSerLeuAlaGaspSerGlyVal    201
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
OY     394   ACCGTGTTTCATTGTGACGACCGGCGGAGCAACTTCTCGAGAGCTGTACGCGCTGCTCA    453
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     202   GluIlePheThrPheGlyIleTyrGlnGlyAsnIleArgGluLeuAlaAspMetaIaSer    221
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
OY     454   GCCCTGCGCAGAGACACTGCACCTTGTG    483
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     222   ThrProLySGluGlnHisCytryLeuIeu    231
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 13
US-10-453-372-180
; Sequence 180, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alcobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187

```

```

      / PRIOR FILING DATE: 2001-03-29
      / PRIOR APPLICATION NUMBER: 60/195792
      / PRIOR FILING DATE: 2000-03-10
      / PRIOR APPLICATION NUMBER: 09/839446
      / PRIOR FILING DATE: 2001-03-19
      / PRIOR APPLICATION NUMBER: 60/199476
      / PRIOR FILING DATE: 2000-03-25
      / PRIOR APPLICATION NUMBER: 09/863776
      / PRIOR FILING DATE: 2001-05-23
      / PRIOR APPLICATION NUMBER: 60/208263
      / PRIOR FILING DATE: 2000-05-31
      / PRIOR APPLICATION NUMBER: 09/939398
      / PRIOR FILING DATE: 2001-08-24
      / PRIOR APPLICATION NUMBER: 60/227800
      / PRIOR FILING DATE: 2000-08-25
      / Remaining Prior Application data removed - See File Wrapper or PALM.
      / NUMBER OF SEQ ID NOS: 1609
      / SOFTWARE: CurnaseqList version 0.1
      / SEQ ID NO 180
      / LENGTH: 709
      / TYPE: PRT
      / ORGANISM: Homo sapiens
US-10-453-372-180

Alignment Scores:
Pred.:          1.25e+05           Length:       709
Score:          161.00             Matches:      53
Percent Similarity: 46.7%         Conservative: 26
Best Local Similarity: 31.4%       Mismatches:   76
Query Match:     15.3%            Indels:       14
DB:              6                Gaps:         4

US-10-699-035A-1 (1-537) x US-10-453-372-180 (1-709)

QY      4 GACCTGAATGTTCTCCTGTGGACAGCTCAGCCAGCGTCTCACTACGATTCTCCCGGATT 63
        ::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db      68 GlusEuValPheLeuValAlaAspSerSerSerValGlglValAlaSnHeArgSerGu 87
        |||||:::|||:::|||:::|||:::|||:::|||:::|||

QY      64 CGGAGATTGTGGGGAGCTGTGGCTCCACTGCCTCCCGGACCAGCGGCCCTGCTGCC 123
        ::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db      88 LeuMetPheValArgLysLeuLeuSerAerPheProValIprOthrAlaThrArgVal 107
        |||||:::|||:::|||:::|||:::|||:::|||:::|||

QY      124 AGCTGTGTGACAGTGAGGACAGTGGGCCATTCACCGAGTCCCTTCCGCCACGACAGCTCG 183
        :|||||:::|||:::|||:::|||:::|||:::|||:::|||
Db      108 AlaIleValThrPheSerSerLysAsnTyValValProArgValAspTryrIleSerThn 127
        :|||||:::|||:::|||:::|||:::|||:::|||:::|||

QY      184 GGTAGGAGCTGCCAG-----GATGCGGTGGCTGCTTCTGCCAG 222
        |||||:::|||:::|||:::|||:::|||:::|||:::|||
Db      128 ArgAlaGlaIahArgGlnHisLysCysAlaLeuLeuLeuGlnGluIleProAlaIleSerTy 147
        |||||:::|||:::|||:::|||:::|||:::|||:::|||

QY      223 CGCATGGGTGACACCCACACTGGCTGGCGCTGCTGTATGCCAAGAACAGCTGTTGCT 282
        |||||:::|||:::|||:::|||:::|||:::|||:::|||
Db      148 ArgGlglglglgltHrThyTrHrLysglAlaPheGlnGlnAlaIaGlnIleLeuLeu--- 166
        |||||:::|||:::|||:::|||:::|||:::|||:::|||

QY      283 GAAGACATCAAGTCCCCGCCAGAGGGGTGCCCAAAGTCTGTGGGTGACAGATGGCGGC 342
        |||||:::|||:::|||:::|||:::|||:::|||:::|||
Db      167 -----HisAlaArgGluAsnSerThrLysValValPheLeuIleThrAspGlyTy 183
        |||||:::|||:::|||:::|||:::|||:::|||:::|||

QY      343 TCCAGC-----GACCCTGTGGGGCCCCCACAGCAGGAGCTCAAGACCTGGGGCTGCACC 396
        |||||:::|||:::|||:::|||:::|||:::|||:::|||
Db      184 SerAsnGlglglAspTro--ArgProIleAlaIaSerLeuLysAspSerGlglValGlu 202
        |||||:::|||:::|||:::|||:::|||:::|||:::|||

QY      397 GTGTTCATTGTGACAGCCGCCAGGAGCAATTCTTGAGCTGTCAACCGCTGCTCAGCC 456
        ::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db      203 ILePheThrPheGlylLetrpgInglYanIlleArgGluLeuAsnAspMetAlaSerThn 222
        ::|||:::|||:::|||:::|||:::|||:::|||:::|||

QY      457 CCTGCCGAGACACCTGCACTTTGTG 483
        |||:::|||:::|||:::|||:::|||:::|||:::|||
Db      223 ProLysGluGluHisCysTyTrLeuLeu 231
        |||:::|||:::|||:::|||:::|||:::|||:::|||

RESULT 14
US-10-453-372-194
/ Sequence 194, Application US/10453372
```

```
; Publication No. US2006000323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 194
; LENGTH: 3568
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-194

Alignment Scores:
Pred. No.: 1,14e-05 Length: 3568
Score: 161.00 Matches: 53
Percent Similarity: 46.7% Conservative: 26
Best Local Similarity: 31.4% Mismatches: 76
Query Match: 15.3% Indels: 14
DB: Gaps: 4

US-10-699-035A-1 (1-537) x US-10-453-372-194 (1-3568)
QY 4 GACCTGATGTTCTCTGCTGACAGCTCAGCCAGCTCTCTACTAGAGTTCTCCGGGTT 63
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
80 GluLeuValPheLeuValAspAspSerSerValGlyGluValAsnPhaArgSerGlu 99
QY 64 CGGAGATTGTTGGGGCAGCTGGTGGCTCCACTGCCCCCTGGGACCGGGGCTGGTGCC 123
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
100 LeuMetPheValArgGlyLeuLeuSerAspPheProValValProThrAlaThrArgVal 119
QY 124 AGCTGTGTCAGCTGGGCGAGTCGCCATACACCGAGTTCCCTTGGCCAGACAGCTCG 183
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
120 AlaIleValThrPheSerSerIysAsnTyValValProArgValAspTyrIleSerThr 139
QY 184 GGTGAGGCTGCCAG-----GATGGGTGGCTGCTTCTGCCCCAG 222
Db 140 ArgArgAlaArgGlnHisIleLysCysAlaLeuLeuLeuGlnGlnIleProAlaIleSerTyr 159
QY 223 CGCATGTGGTACACCACTGAGCTGGCGCTGCTATGCAAGAAAGAGCTGTTGTT 282
Db 160 ArgGlyGlyGlyThrTyrThrTyrGlyAlaPheGlnGlnAlaGlnIleLeuLeu--- 178
QY 283 GAAGCATCAGGTGCTCCCGCCAGGGGTGCCAAAGTCTGTGGGTGACAGATGGCGGC 342
Db 179 -----HisAlaArgGlnHisSerThrLysValValPheLeuIleThrAspGlyTyr 195
QY 343 TTCAGC-----GACCTGTGGGCCCCCCCATGACGAGAGCTCAAGAGCTGGGCGTCC 396
Db 196 SerAsnGlyGlyAspPro---ArgProIleAlaAlaSerLeuArgAspSerGlyValGln 214
```

```
QY 397 GTGTTCAATGTCAGCACCGGCGAGCACTTCTGAGCTGTCTACGCGCTCCAGCC 456
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
215 IlePheThrPheGlyIleIleTyrGlnIleAsnIleArgGlnLeuAsnAspMetAlaSerThr 234
QY 457 CTGCGCGAAGACACCTTGACTTGTG 483
Db 225 ProLysGlnGlnHisCysTyrLeuLeu 243

RESULT 15
US-10-453-372-178
; Sequence 178, Application US/10453372
; Publication No. US2006000323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 178
; LENGTH: 3570
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-178

Alignment Scores:
Pred. No.: 1,14e-05 Length: 3570
Score: 161.00 Matches: 53
Percent Similarity: 46.7% Conservative: 26
Best Local Similarity: 31.4% Mismatches: 76
Query Match: 15.3% Indels: 14
DB: Gaps: 4

US-10-699-035A-1 (1-537) x US-10-453-372-178 (1-3570)
QY 4 GACCTGATGTTCTCTGCTGACAGCTCAGCCAGCTCTCTACTAGAGTTCTCCGGGTT 63
Db 80 GluLeuValPheLeuValAspAspSerSerValGlyGluValAsnPhaArgSerGlu 99
QY 64 CGGAGATTGTTGGGGCAGCTGGTGGCTCCACTGCCCCCTGGGACCGGGGCTGGTGCC 123
Db 100 LeuMetPheValArgGlyLeuLeuSerAspPheProValValProThrAlaThrArgVal 119
QY 124 AGCTGTGTCAGCTGGGCGAGTCGCCATACACCGAGTTCCCTTGGCCAGACAGCTCG 183
Db 120 AlaIleValThrPheSerSerIysAsnTyValValProArgValAspTyrIleSerThr 139
QY 184 GGTGAGGCTGCCAG-----GATGGGTGGCTGCTTCTGCCCCAG 222
Db 140 ArgArgAlaArgGlnHisIleLysCysAlaLeuLeuLeuGlnGlnIleProAlaIleSerTyr 159
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 13:10:29 ; Search time 9.08696 Seconds
(without alignments)
603.637 Million cell updates/sec

Title: US-10-699-035A-6

Perfect score: 2154
Sequence: 1 MLPWTALGLALSLRLALARS.....RRPPRPVAPAPPGTASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 1312538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA_New:*
1: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB pep:*
2: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB pep:*
3: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB pep:*
4: /cgn2_6/prodata/2/pubppaa/PCT_NEW_PUB pep:*
5: /cgn2_6/prodata/2/pubppaa/US09_NEW_PUB pep:*
6: /cgn2_6/prodata/2/pubppaa/US10_NEW_PUB pep:*
7: /cgn2_6/prodata/2/pubppaa/US11_NEW_PUB pep:*
8: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2130.5	98.9	445	US-10-453-372-2	Sequence 2, App11
2	421.5	19.6	3063	US-11-186-284-26	Sequence 26, App1
3	401	18.6	517	US-11-169-041-160	Sequence 160, App
4	253.5	11.8	915	US-10-131-826A-294	Sequence 294, App
5	253.5	11.8	956	US-11-113-424-39	Sequence 39, App1
6	194	9.0	214	US-11-192-449-6	Sequence 6, App11
7	194	9.0	214	US-11-192-449-9	Sequence 9, App11
8	185	8.6	214	US-11-192-449-5	Sequence 5, App11
9	184	8.5	678	US-10-063-703-34	Sequence 34, App1
10	184	8.5	678	US-11-102-240-34	Sequence 34, App1
11	175	8.1	709	US-10-453-372-186	Sequence 186, App
12	174	8.1	709	US-10-453-372-180	Sequence 180, App1
13	174	8.1	1152	US-11-080-026-4	Sequence 4, App11
14	173	8.0	709	US-10-453-372-182	Sequence 182, App
15	172.5	8.0	1167	US-10-601-368-18	Sequence 18, App1
16	169	7.8	3570	US-10-453-372-178	Sequence 178, App
17	169	7.8	3570	US-10-453-372-196	Sequence 196, App
18	169	7.8	3570	US-10-453-372-198	Sequence 198, App
19	169	7.8	3570	US-10-453-372-200	Sequence 200, App
20	169	7.8	3570	US-10-453-372-202	Sequence 202, App
21	169	7.8	3570	US-10-453-372-204	Sequence 204, App
22	168	7.8	709	US-10-453-372-184	Sequence 184, App
23	168	7.8	709	US-10-453-372-184	Sequence 184, App
24	168	7.8	3568	US-10-453-372-194	Sequence 194, App
25	162.5	7.5	847	US-10-995-561-634	Sequence 634, App

26	162.5	7.5	1259	6	US-10-995-561-625	Sequence 625, App
27	162.5	7.5	1286	6	US-10-995-561-628	Sequence 628, App
28	162.5	7.5	1341	6	US-10-995-561-621	Sequence 621, App
29	162.5	7.5	2217	7	US-11-193-561-38	Sequence 38, App1
30	162.5	7.5	2217	7	US-11-193-561-38	Sequence 38, App1
31	162.5	7.5	2217	7	US-11-193-789-38	Sequence 38, App1
32	162.5	7.5	2217	7	US-11-193-806-38	Sequence 38, App1
33	162.5	7.5	2217	7	US-11-193-857-38	Sequence 38, App1
34	162.5	7.5	2330	7	US-11-193-561-21	Sequence 21, App1
35	162.5	7.5	2330	7	US-11-193-771-21	Sequence 21, App1
36	162.5	7.5	2330	7	US-11-193-789-21	Sequence 21, App1
37	162.5	7.5	2330	7	US-11-193-806-21	Sequence 21, App1
38	162.5	7.5	2330	7	US-11-193-857-21	Sequence 21, App1
39	162.5	7.5	2355	6	US-10-995-561-623	Sequence 623, App
40	162.5	7.5	2355	6	US-10-995-561-627	Sequence 627, App
41	162.5	7.5	2355	7	US-11-193-561-19	Sequence 19, App1
42	162.5	7.5	2355	7	US-11-193-771-19	Sequence 19, App1
43	162.5	7.5	2355	7	US-11-193-789-19	Sequence 19, App1
44	162.5	7.5	2355	7	US-11-193-806-19	Sequence 19, App1
45	162.5	7.5	2355	7	US-11-193-857-19	Sequence 19, App1

ALIGNMENTS

RESULT 1
US-10-453-372-2
; Sequence 2, Application US/10453372
; Publication No. US20060003323A1
GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 2
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-2
Query Match 98.9%; Score 2130.5; DB 6; Length 445;
Best Local Similarity 93.9%; Pred. No. 3.9e-161;
Matches 418; Conservative 0; Mismatches 0; Indels 27; Gaps 1;
QY 1 MLPWTALGLALSLRLALARS...RRPPRPVAPAPPGTASREP 418
Db 1 MLPWTALGLALSLRLALARS...RRPPRPVAPAPPGTASREP 418
QY 61 VAPLPGLTALRASLVHVSRRPYTEPPFGQSSGGEAQAQVAPASQRMGDTHTGLALVYA 120

```
Db      61 VAPLPLGTALRASLHVSGSRPYTERPFQOHSSGEAQAQDVRASQRMGDTHGLALVYA 120
Qy      121 KEOLFASAGARPVGVKVLVWTDGSSDPVGPMPQELDGLVTVFIVSTGRGNFLELSA 180
Db      121 KQOLFASAGARPVGVKVLVWTDGSSDPVGPMPQELDGLVTVFIVSTGRGNFLELSA 180
Qy      181 MASAPAEKHLHFVDVDDLHIIYOELRGSILDMARPOOLHATEITSSGFLAMPPLITADS 240
Db      181 MASAPAEKHLHFVDVDDLHIIYOELRGSILDMARPOOLHATEITSSGFLAMPPLITADS 240
Qy      241 GYVVELVPSAOPGAAARQQLPQNATDWIMAGLDPTDVALVPSNVRLRPOILRR 300
Db      241 GYVVELVPSAOPGAAARQQLPQNATDWIMAGLDPTDVALVPSNVRLRPOILRR 300
Qy      301 TR-----PEAGPERIVISHARPSRLVSNAPALGSAA 333
Db      301 TRPGEAGPGASGPESGAGAPPTQALAPAEAGPERIVISHARPSRLVSNAPALGSAA 360
Qy      334 ALGYHVQFGLRNGEAGQVVEVPAGRNCTTLOGIAPGTAVLVVTAAFRSGRESALSAAK 393
Db      361 ALGYHVQFGLRNGEAGQVVEVPAGRNCTTLOGIAPGTAVLVVTAAFRSGRESALSAAK 420
Qy      394 TPDGPRPRRPVPRAPTGTASREP 418
Db      421 TPDGPRPRRPVPRAPTGTASREP 445
```

```
RESULT 2
US-11-186-284-26
; Sequence 26, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Millenium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPW01-029P2RM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3063
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-26
```

```
Query Match      19.6%; Score 421.5; DB 7; Length 3063;
Best Local Similarity 27.5%; Pred. No. 5.5e-25;
Matches 128; Conservative 59; Mismatches 174; Indels 105; Gaps 9;
```

```
Db      498 TKVEDIIEAINTFPYRGSSTNTGKMTYVREKIFVPSKGSNSNPKMILITDGSSDAF 557
Qy      152 GPPMOELDGLVTVFIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIYOELRGS1- 209
Db      558 RDPALKRNSDVEIFAVGVKAQVAVRSELEIASPPAEHTVFTVEDDAPORISFELTQSLC 617
Qy      210 -----LDAMR-----POOLHATEITSSGFLAMPPL-----LTADSGIVV 244
Db      618 LRIEOLAAIKKAVVPPKDLSEFSEVTSYGFKTNMSPAGENVSYHITYKEAAGDEYTV 677
Qy      245 LE-----LVPSAOP-----GAAARQQLPQNATD 267
Db      678 VEPASTSVLSLKPETILVNVTAIEYEDGRSIFLAGEETTEEKGAAPRNLKATDETDT 737
Qy      268 -----WMA-----GLDPTDVALVPSN 288
Db      738 SFKITWQAQGRVLRCLIIYRPVAGSEREVTPPNQRRRTLENLIPDKVEVSIVPEVF 797
Qy      289 VRLAPQILRVKTRPEEAGPERIVISHARPSRLVSNAPALGSAAALGYHVQFPLRGCE 348
Db      798 SGPGTPLTGMNATEBEVGNPDLRVSDPTTSTMKLSWGAPGKVKQ--YLVTYTPVAGGE 855
Qy      349 AQRVEVPAGRNCTTLOGIAPGTAVLVVTAAFRSGRESALSAAKACT 394
Db      856 TOEVTVRGDTTNTVLOGLKEGTQYALSVTALYASAGDALRGEGTT 901
```

```
RESULT 3
US-11-169-041-160
; Sequence 160, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 160
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-169-041-160
```

```
Query Match      18.6%; Score 401; DB 7; Length 517;
Best Local Similarity 30.0%; Pred. No. 2.2e-24;
Matches 114; Conservative 61; Mismatches 179; Indels 26; Gaps 6;
```

QY 318 PRSLRVSAPALGSAALGYHVOFGPRLRGSAQRYEVPAGNCTTLOGIAPGTAVLVTVT 377
DB 437 TDSFRTSTQAPGK--VLRVRIYRPVAGGESREVTTPNQRRITLLENLIPOTKEVSIVT 494
QY 378 AAFRSGRESAISAKACTPDG 397
DB 495 PEYFSGPGTPLTGNMATEBG 514

RESULT 4
US-10-131-826A-294
; Sequence 294, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C128
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056994
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PAM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 294
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-294

Query Match 11.8%; Score 253.5; DB 6; Length 915;
Best Local Similarity 31.6%; Pred. No. 2.1e-12;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERGPPASARGDLMFLDSSASVSHYEFGRVREFVQGVAPLPLGTGALRASL 75
DB 45 ALLESSCEN-----KRADLVFTIDSSRSVNTHDYAKVKEFIVDIIQFLDIGDVTVRGVL 98

QY 76 VHVGRPYTEPPFPGHSSGEAAQDAVRASAOQMDPTHGLLVAKEOLEFAEASGARP-- 133
DB 99 LQYGSTVKNESLTKTKRKSEVERAVKMRHLSTGTMTGLAIQYALNTAFSABGARPLR 158
QY 134 -GVPKVLVWVTDGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLTAASAASAPAEKHL-- 190
DB 159 ENVPKRVIMVTDGRPDQDSVAEVAARAKARDTGILIFAIQGVQVDFNTLKSIGSEPHDHVFL 218
QY 191 --HFVDVDDLHIIVQE 204
DB 219 VANFSQIETLTSVFOK 234

RESULT 5
US-11-113-424-39
; Sequence 39, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangoli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-39

Query Match 11.8%; Score 253.5; DB 7; Length 956;
Best Local Similarity 31.6%; Pred. No. 2.2e-12;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERGPPASARGDLMFLDSSASVSHYEFGRVREFVQGVAPLPLGTGALRASL 75
DB 45 ALLESSCEN-----KRADLVFTIDSSRSVNTHDYAKVKEFIVDIIQFLDIGDVTVRGVL 98
QY 76 VHVGRPYTEPPFPGHSSGEAAQDAVRASAOQMDPTHGLLVAKEOLEFAEASGARP-- 133
DB 99 LQYGSTVKNESLTKTKRKSEVERAVKMRHLSTGTMTGLAIQYALNTAFSABGARPLR 158
QY 134 -GVPKVLVWVTDGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLTAASAASAPAEKHL-- 190
DB 159 ENVPKRVIMVTDGRPDQDSVAEVAARAKARDTGILIFAIQGVQVDFNTLKSIGSEPHDHVFL 218
QY 191 --HFVDVDDLHIIVQE 204
DB 219 VANFSQIETLTSVFOK 234

RESULT 6
US-11-192-449-6
; Sequence 6, Application US/11192449

```

; Publication No. US2005028181A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USCn3
; CURRENT APPLICATION NUMBER: US/11/192,449
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
; US-11-192-449-6

```

Query Match	9.0%;	Score 194;	DB 7;	Length 214;
Best Local Similarity	33.5%;	Pred. No. 1.6e-08;		
Matches 63;	Conservative 28;	Mismatches 77;	Indels 20;	Gaps 7

Qy	34	DLMFLDSSASVSHSEFSVREFFVQLVPLDLGICALASLVHGSRYTEFFPQGHSS	93
Db	22	DIVIVLDGSSNLT--YPMDSVTAFLNDLKRMLIGPKQTVGGLVQGENANTHEFNANKYSS	79
Qy	94	GEAADAARASAOQRMG-DTHHTGLALVYAKEQLFAEASGARPGVPRVLVWVTDGSSD--P	150
Db	80	TEEVLVAAKKITVQRGSRQTMETALGTDTPARKAEAFTEARGARPGVKMVLVTGDESHDNR	139
Qy	151	VGPPEQELKDLGVTFIV---STGRGNLT-----ELSAASAPAEKHLHFVVDVDD---	197
Db	140	LKKVIQDCEDENIQRFSIALIGSSYNKGNLSTEKFVEEIKSIASEPTEK--HEFNVSDELA	197
Qy	198	LHIIVQEL 205	
Db	198	LVTIVKTL 205	

```

, RESULT 7
, US-11-192-449-9
, Sequence 9, Application US/11192449
, Publication No. US20050281818A1
, GENERAL INFORMATION:
, APPLICANT: Biogen, Inc.
, APPLICANT: Gotwals, Philip
, APPLICANT: Koteliansky, Victor
, TITLE OF INVENTION: Method for the Treatment of Fibrosis
, FILE REFERENCE: A073-USCN3
, CURRENT APPLICATION NUMBER: US/11/192,449
, CURRENT FILING DATE: 2005-07-28
, PRIOR APPLICATION NUMBER: 10/625,260
, PRIOR FILING DATE: 2003-07-22
, PRIOR APPLICATION NUMBER: 10/061,658
, PRIOR FILING DATE: 2002-02-01
, PRIOR APPLICATION NUMBER: 09/557,092
, PRIOR FILING DATE: 2000-04-21
, PRIOR APPLICATION NUMBER: 60/130,847
, PRIOR FILING DATE: 1999-04-22
, PRIOR APPLICATION NUMBER: 60/137,214
, PRIOR FILING DATE: 1999-06-01
, NUMBER OF SEQ ID NOS: 10
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 9

```

```

; LENGTH: 214
;
; TYPE: PRT
; ORGANISM: Homo sapien
US-11-192-449-9

```

Query Match	9.0%;	Score 194;	DB 7;	Length 214;
Best Local Similarity	33.5%;	Pred. No. 1.6e-08;		
Matches 63;	Conservative 28;	Mismatches 77;	Indels 20;	Gaps 7;

```
Oy      34ADLMFLLSSASVSHYEFRRVRFEFGQLAFLPGLTGALRASLIYGVSRPYTEFFGQHS   93
       |||:|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      22DIVIVLDSNST--YPMDSVTALLNDLLKRMIDGPKQIQVGITYOYGENTHREFNINKYS   79
```

QY 94 GEAADNVAASQRMG-DTHGTALVAKEQLPAEASGRPGYKVLVWVTDGSSD--P 150
 Db 80 TEEVLVAAKKIYGRGRTMTALCTDTPARKEFTFARGARGVKKMVIITDGSHDNHR 139

QY
151 VGGPPEMOELDGLVTFIV---STGRNLT-----EISAAASAPAEKHLHFVVDD---197

DB
140 LKKVIQDCEDENITQRFSAILGSGYNNRGNLSTEKEVEELSIASISPTKE--HFFNVSDLLA 197

QY 198 LHI VOEL 205
| | | |
| | : |
DB 198 LVTIVKTL 205

RESULT 8

; Sequence 5, Application US/11192449
; Publication No. US20050281818A1

```

1  APPLICANT: Biogen, Inc.
2  APPLICANT: Gotwals, Philip
3  APPLICANT: Koreliansky, Victor
4  TITLE OF INVENTION: Method for the Treatment of Fibrosis
5  FILE REFERENCE: A073-USCN3
6  CURRENT APPLICATION NUMBER: US/11/192,449
7  CURRENT FILING DATE: 2005-07-28
8  PRIOR APPLICATION NUMBER: 10/625,260
9  PRIOR FILING DATE: 2003-07-22
10 PRIOR APPLICATION NUMBER: 10/061,658
11 PRIOR FILING DATE: 2002-02-01
12 PRIOR APPLICATION NUMBER: 09/557,092
13 PRIOR FILING DATE: 2000-04-21
14 PRIOR APPLICATION NUMBER: 60/130,847
15 PRIOR FILING DATE: 1999-04-22
16 PRIOR APPLICATION NUMBER: 60/137,214
17 PRIOR FILING DATE: 1999-06-01
18
19

```

```

: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5
:
: LENGTH: 214
:
: TYPE: PRT
: ORGANISM: Rat
: US-11-192-449-5

```

Query Match	8.6%;	Score 185;	DB 7;	Length 214;
Best Local Similarity	31.9%;	Pred. No. 8e-08;		
Matches 60;	Conservative 32;	Mismatches 76;	Indels 20;	Gaps 7;

Qy 34DLNLFSSASVSHYEFGRVRREFGQLVARPLPTGTALRASLIYHGSRYTEPFPGOHSS 93
::
Db 22DIYIVLDSNST--YPWESVIAPFLNDLKRMIDICPKQTGVIGVOYGENTHEFNINLKYS 79

Qy 94GEAQAQDAVASAQMVG-DTHGTGLALVNAEQLFEEASGARPGVSKVLVWVTDGSSD--P 150
 80 TEEVLVAAKKIGKGGLOTMTALGIDTARKEAFIARGARGVKVMVIVTDGSHDNYR 139
 Db

Oy 151 VCGMOELKDLGTVFIVST-----GRGNL-----ELSAASAAPAEKHILHFVDD-- 197
Db 140 LKVQIQDCEDENIORFSIALGLHYNRGNLSTEKEVEEIRISIASETEK--HFFNVSDLELA 197

QY 198 LHIIVQEL 205

QY 198 LHIIVQEL 205


```
QY 7 LGLALSLRLALASGAEGRPPASAPRGDLMFLDSSASVSHYFESRVEFPVQLVAPLPL 66
DB 49 LGOAFRRRVRLLREJSE-----RLLEIVFLVDDSSSVGEVNFSELMFVVKLLSDPFV 100
QY 67 GTGALRASLVHVGSRPYTEFPFGQSSGSAQ-----DAVRASQMGDTHGLALV 118
DB 101 VFTATRVAVLTSSKNYV-VPRADYISTRARQHKCALLQETPAISYRGGTYTKGAQ 159
QY 119 YAKEQLFAEASGAPGVPKVLVWVTDGSS--DPVGPMEQLKDLGVTFIVSTGRNFL 176
DB 160 QAAQILL-----HARENSAKVFLITDGYNGDP-RPIASLRDSGVEIFTGIMQGNIR 214
QY 177 ELISAASAPAEKHLHFVDDVLDHIIYQELRGSLDAMRPOQLHATITSSGF 228
DB 215 ELNDMASTPEKEHCYL-----LHSFEE-----FEALARRALH-BDLPESGSF 254

RESULT 12
US-10-453-372-180
; Sequence 180, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT FILING DATE: 2003-06-03
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 180
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-180

Query Match 8.1%; Score 174; DB 6; Length 709;
Best Local Similarity 29.4%; Pred. No. 2.9e-06;
Matches 68; Conservative 33; Mismatches 96; Indels 34; Gaps 8;

QY 7 LGLALSLRLALASGAEGRPPASAPRGDLMFLDSSASVSHYFESRVEFPVQLVAPLPL 66
DB 49 LGOAFRRRVRLLREJSE-----RLLEIVFLVDDSSSVGEVNFSELMFVVKLLSDPFV 100
QY 67 GTGALRASLVHVGSRPYTEFPFGQSSGSAQ-----DAVRASQMGDTHGLALV 119
DB 101 VFTATRVAVLTSSKNYVPRVDYISTRARQHKCALLQETPAISYRGGTYTKGAQ 160
QY 120 YAKEQLFAEASGAPGVPKVLVWVTDGSS--DPVGPMEQLKDLGVTFIVSTGRNFL 177
DB 161 QAAQILL-----HARENSAKVFLITDGYNGDP-RPIASLRDSGVEIFTGIMQGNIR 215
QY 178 LSAAASAPAEKHLHFVDDVLDHIIYQELRGSLDAMRPOQLHATITSSGF 228
```

```
DB 216 LNDMASTPEKEHCYL-----LHSFEE-----FEALARRALH-BDLPESGSF 254

RESULT 13
US-11-080-026-4
; Sequence 4, Application US/11080026
; Publication No. US20050260192A1
; GENERAL INFORMATION:
; APPLICANT: Springer, Timothy A.
; APPLICANT: Shimaoka, Motomu
; APPLICANT: Lu, Chafen
; TITLE OF INVENTION: MODIFIED POLYPEPTIDES STABILIZED IN A
; FILE REFERENCE: CFBF-P02-021
; CURRENT FILING DATE: US/11/080,026
; PRIOR APPLICATION NUMBER: 2005-03-15
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/229,700
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1152
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-080-026-4

Query Match 8.1%; Score 174; DB 7; Length 1152;
Best Local Similarity 25.0%; Pred. No. 5.5e-06;
Matches 74; Conservative 42; Mismatches 122; Indels 58; Gaps 11;

QY 2 LPTALGLALSLRLALASGAE-----RGPPASAP- 31
DB 80 VVEAVNMSLGLSLAATTSPPQLACGPTVHOTCSENTYVKGCLFPGSNLRQOQKPE 139
QY 32 --RG-----DIMFLDSSASVSHYFESRVEFPVQLVAPLPLGTGALRASLVHVSRLPY 83
DB 140 ALRGCPQEDSDIAPLIDSGSITIPHDPRMKEFVSTVMEQLKSKYTLF--SIMQYSEBR 197
QY 84 TFFPFGQSSGSAADDAVRASQMGDTHGLALVYAEQQLFAEASGAPGVPKVLVWVT 143
DB 198 IHFTFKEPQNNPNPSVLVPIITQLGRTHATGIRKVVRELFINTGARKNAFKILLVLT 257
QY 144 DGGS--SDPVG--PMEQLKDLGVTFIVSTG-----RGNFELSAASAPAEKHLHFV-D 194
DB 258 DEKRGDPLGYEDVLPEDAREGVIRYICVGDAFRSEKSRQELNITASKPRPDHYFOVNN 317
QY 195 VDDLHIIYQELRGSLDAMRPOQLHAT-----ETSSGFRLLAMP--PLITADSGY 242
DB 318 PALKTIQNLREKIF-AIEGTGTGSSSSEFHEHMQEGPSAITSNGPLSTVGSY 372

RESULT 14
US-10-453-372-182
; Sequence 182, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT FILING DATE: US/10/453,372
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
```


THIS PAGE BLANK (USPTO)

Db 92 SSGEAAQAVRASAGRMEDTHTGLALVYAKEDLFEASGARGVPRVLTWVTDGSSDPV 151
QY 121 GPPMOELKDLGTVTVIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
Db 152 GPPMOELKDLGTVTVIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211

RESULT 2

US-11-186-284-26

Sequence 26, Application US/11186284
Publication No. US20050266493A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF COLON CANCER
FILE REFERENCE: MEM01-029F22RNM
CURRENT APPLICATION NUMBER: US/11/186,284
CURRENT FILING DATE: 2005-07-21
PRIOR APPLICATION NUMBER: US/10/301,822
PRIOR FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 26
LENGTH: 3063
TYPE: PRT
ORGANISM: Homo Sapiens
US-11-186-284-26

Query Match 28.9%; Score 263.5; DB 7; Length 3063;
Best Local Similarity 35.8%; Pred. No. 1.1e-17;
Matches 64; Conservative 30; Mismatches 84; Indels 1; Gaps 1;

QY 1 RGLDMFLDSSASVSHYEFRRVRFVQGLVAPLPLGTGALRASLVHVGSRPTTFFPFGH 60
Db 438 KADIVFLVDGSSISIGIANFVKRAFLVLYKSFELSPNRVQISLVQYSRDPTEFTLKKF 497
QY 61 SSGEAAQAVRASAGRMEDTHTGLALVYAKEDLFEASGARGVPRVLTWVTDGSSDPV 120
Db 498 TKVEIIEIINTFPFRGSGTNGKAMTYREKIFVPSKGSRSNVKWMILITDGSSDAF 557
QY 121 GPPMOELKDLGTVTVIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGSI 178
Db 558 RDPALKNSDVEIFAVGVKDAVSELEAIASPPAETHVFTVEDFADFQRISEFELTQSI 616

RESULT 3

US-11-169-041-160

Sequence 160, Application US/11169041
Publication No. US20060019284A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
FILE REFERENCE: 10001 NP
CURRENT APPLICATION NUMBER: US/11/169,041
CURRENT FILING DATE: 2005-06-28

QY 5 MFLDSSASVSHYEFRRVRFVQGLVAPLPLGTGALRASLVHVGSRPTTFFPFGHSSGE 64
Db 143 MFLVDGSSISIGIANFVKRAFLVLYKSFELSPNRVQISLVQYSRDPTEFTLKKF 202
QY 65 AAQDAVRASAGRMEDTHTGLALVYAKEDLFEASGARGVPRVLTWVTDGSSDPVGP 124
Db 203 DIIEMINTFPFRGSGTNGKAMTYREKIFVPSKGSRSNVKWMILITDGKSSDAFRDPA 262
QY 125 QELMDLGTVTVIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGSI 178
Db 263 IKLRNSDVEIFAVGVKDAVSELEAIASPPAETHVFTVEDFADFQRISEFELTQSI 317

Query Match 28.2%; Score 257.5; DB 7; Length 517;
Best Local Similarity 36.6%; Pred. No. 4.1e-18;
Matches 64; Conservative 27; Mismatches 83; Indels 1; Gaps 1;

US-11-169-041-160

Sequence 294, Application US/10131826A
Publication No. US20050245730A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroft, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gunney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C128
CURRENT APPLICATION NUMBER: US/10/131,826A
CURRENT FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588

```

; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 294
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-294

Query Match      27.5%; Score 251.5; DB 6; Length 915;
Best Local Similarity 32.2%; Pred. No. 3.5e-17;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

QY 1 RGDLMFLDSSASVSHYEFRRVEYQGVAPLPCTGALRASLVHSGRPYTEPPFGH 60
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 55 RADLVFIIDSSRSVNHDAKYAEFIVDILQFLDIGPDVTRVGLQYGSSTVKNESLKTFF 114
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 61 SSGEAAQDVRASAGMGTHTGLALVYAKEQLFANASGARP---GVPRVLVWMTDGGSS 117
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 115 KRKSEVERAVKRMRLSTGTMTGLAQYALNIAFSSAEAGARPLRENVPRVIMIVTDGRPO 174
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 118 DPVGPMPQELKDLGVTVFIVSTGRGNFLLSAASAPAEKHL---HFVDVDDLHIYOE 173
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 175 DSVAEVAAKARDTGILIFAIIGVQVDFNTLKSIGSEPHEDHVFIVANFSQIETLTSVPQK 234
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 5
US-11-113-424-39
; Sequence 39, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangoli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; PRIOR FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 39
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-39

Query Match      27.5%; Score 251.5; DB 7; Length 956;
Best Local Similarity 32.2%; Pred. No. 3.7e-17;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

QY 1 RGDLMFLDSSASVSHYEFRRVEYQGVAPLPCTGALRASLVHSGRPYTEPPFGH 60
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 55 RADLVFIIDSSRSVNHDAKYAEFIVDILQFLDIGPDVTRVGLQYGSSTVKNESLKTFF 114
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 61 SSGEAAQDVRASAGMGTHTGLALVYAKEQLFANASGARP---GVPRVLVWMTDGGSS 117
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 115 KRKSEVERAVKRMRLSTGTMTGLAQYALNIAFSSAEAGARPLRENVPRVIMIVTDGRPO 174
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

```

QY 118 DPVGPMPQELKDLGVTVFIVSTGRGNFLLSAASAPAEKHL---HFVDVDDLHIYOE 173
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 175 DSVAEVAAKARDTGILIFAIIGVQVDFNTLKSIGSEPHEDHVFIVANFSQIETLTSVPQK 234
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 6
US-11-192-449-6
; Sequence 6, Application US/11192449
; Publication No. US20050281818A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliensky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USC93
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/557,092
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-11-192-449-6

Query Match      21.2%; Score 194; DB 7; Length 214;
Best Local Similarity 33.5%; Pred. No. 2.7e-12;
Matches 63; Conservative 28; Mismatches 77; Indels 20; Gaps 7;

QY 3 DLMFLDSSASVSHYEFRRVEYQGVAPLPCTGALRASLVHSGRPYTEPPFGHSS 62
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 22 DIVIVLDGSSNST--PFWMSVTAFLNDLKRMDIGKQOVGVQYGENVTHEFNINRYS 79
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 63 CEAAQDARASAGMGTHTGLALVYAKEQLFANASGARPVPLVWMTDGGSSD--P 119
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 80 TEEVVAAKKIVQGRGRTMTALGTDPAKCAFTARAGKGVKVMYIVTDGESHDR 139
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 120 VGPPMQELKDLGVTVFIV---STGRGNFL-----ELSAASAPAEKHLHFVDVDD--- 166
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 140 LKVIQDCEDENIQFSAIIGSVYRGNLSTKEFVEIKSIASEBTEK--HFFNVSDELA 197
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 167 LHIIYQEL 174
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 198 LVTIVKTL 205
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 7
US-11-192-449-9
; Sequence 9, Application US/11192449
; Publication No. US20050281818A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliensky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073-USC93
; CURRENT APPLICATION NUMBER: US/11/192,449
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: 10/625,260
; PRIOR FILING DATE: 2003-07-22
; PRIOR APPLICATION NUMBER: 10/061,658
; PRIOR FILING DATE: 2002-02-01
```

```

: PRIOR APPLICATION NUMBER: 09/557,092
: PRIOR FILING DATE: 2000-04-21
: PRIOR APPLICATION NUMBER: 60/130,847
: PRIOR FILING DATE: 1999-04-22
: PRIOR APPLICATION NUMBER: 60/137,214
: PRIOR FILING DATE: 1999-06-01
: NUMBER OF SEQ IDS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 214
: TYPE: prt
: ORGANISM: Homo sapien
US-11-192-449-9

```

Query Match Similarity	21.2%;	Score 194;	DB 7;	Length 214;
Best Local Similarity	33.5%;	Pred. NO. 2.7e-12;		
Matches 63; Conservative	28;	Mismatches 77;	Indels 20;	Gaps 7;

Qy 3 DLMFLDSDASVSHVEFSRVREVGQVLAPELCTGATLRSALVHVSRRPTTEPPQCHSS 62
 Db 22 DIVIIVLDGNSNI--YPMDSVTAFLNDLKRMDIGPKQTOGVIOYGENVTHEFNLUKSS 79
 Qy 63 GEAAODAVASAKORMG-DTHTGIALVYAKEQLPAAASGAPGVPKYLVMWTDGSSD--P 119
 Db 80 TBEVLVAACKIIVQGRGRQIMTALGTGTARKEATFARGARRGVKKVMVITVDDESHDNR 139
 Qy 120 VGPPMOELKDLGVTFIV----STGRNFL-----ELSAASGAPAEKHLAFVDVD-- 166
 Db 140 LKQVIQDCEDENIQRSIALISYNNGNLSTEFVEIEIKSIASEPTEK--HFNVSDELA 197
 Qy 167 LHIIVQEL 174
 Db 198 LVTIVKTL 205

RESULT 8
US-11-192-449-5

```

: Sequence 5, Application US/11192449
: Publication No. US20050281818A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Biogen, Inc.
: APPLICANT: Goetzwa19, Philipp
: APPLICANT: Koteliansk1y, Victor
: TITLE OF INVENTION: Method for the Treatment of Fibrosis
: FILE REFERENCE: A073-USCN3
: CURRENT APPLICATION NUMBER: US/11/192,449
: CURRENT FILING DATE: 2005-07-28
: PRIOR APPLICATION NUMBER: 10/625,260
: PRIOR FILING DATE: 2003-07-22
: PRIOR APPLICATION NUMBER: 10/061,658
: PRIOR FILING DATE: 2002-02-01
: PRIOR APPLICATION NUMBER: 09/557,092
: PRIOR FILING DATE: 2000-04-21
: PRIOR APPLICATION NUMBER: 60/130,847
: PRIOR FILING DATE: 1999-04-22
: PRIOR APPLICATION NUMBER: 60/137,214
: PRIOR FILING DATE: 1999-06-01
:
: NUMBER OF SEQ ID NOS: 10
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 5
:
: LENGTH: 214
:
: TYPE: PRT
:
: ORGANISM: Rat
:
: US-11-192-449-5

```

Query Match	20.3%;	Score 185;	DB 7;	Length 214;
Best Local Similarity	31.9%;	Pred. No. 2.1e-11;		
Matches 60; Conservative	33;	Mismatches 76;	Indels 20;	Gaps 7;

Oy 3 DLMFLDSSASVSHYEFGRVRREFGQLVARPLPGTGLRSLRVHVGSRPYTEFFPFGQHS 62
 |::||| |: : ||: :: |:: |:: |:: |::
Dd 22 DIVIULDGSNST--YPWESVIAPLNDLRKMDIGEQOTGVIVQGENTHEFNLNKYSS 79

QY 63 GEAODAVASARNG -PTHGLAIIVAKKEOLFASASGARPEVPKYLWVTTGGSSD--P 119
 :: ||| ||| |||
Db 80 TEEVLVAKKIKRGROGLOMTALGIDTPARKETFGARGRRGVKKVMVLTDSHDNDR 139
 :: ||| ||| |||
QY 120 VGPRMQLKDLCGYTVFIVST---GRGNFL-----ELSAASAAPAEKLHFVDVD--- 166
 :: ||| ||| |||
Db 140 LKQVIQDEDEDNIQRFTAILGHYNRGNLSTEKFYEIISKISISEPTK--HFFNVSDELA 197
 :: ||| ||| |||
QY 167 LHIIVOEL 174
 ||| |||
Db 198 LVTIIVKAL 205

RESULT 9
US-10-063-703-34

```

; Sequence 34, Application US/10063703
; Publication No. US20060008901A1
; GENERAL INFORMATION:
; APPLICANT: Baton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,703
; PRIORITY FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 34
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-063-703-34

```

Query Match	20.0%	Score 183;	DB 6;	Length 678;
Best Local Similarity	30.6%	Pred. No. 1.6e-10;		
Matches 53; Conservative	25;	Mismatches 91;	Indels 4;	Gaps 2.

[illegible]

RESULT 10
US-11-102-240-34

```

: Sequence 34 Application US/11102240
: Publication No. US20050260647A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: ANTIBODIES TO POLYPEPTIDES ENCODED BY A NUCLEIC ACID UNDEREXPRESSION
: TITLE OF INVENTION: ESOPHAGEAL TUMOR
: FILE REFERENCE: P3230R1C106C
: CURRENT APPLICATION NUMBER: US/11/102,240
: CURRENT FILING DATE: 2005-04-08
: PRIOR APPLICATION NUMBER: 10/063662

```

```

; PRIOR FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: 10/006867
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 60/170262
; PRIOR FILING DATE: 199-12-09
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 34
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-102-240-34
```

```

Query Match          20.0%; Score 183; DB 7; Length 678;
Best Local Similarity 30.6%; Pred. No. 1.6e-10;
Matches 53; Conservative 25; Mismatches 91; Indels 4; Gaps 2;
```

```

QY 3 DLMLFLDSSASVSHYEFSSRVREVGQVLAPLPLGTGALRASLVHVGSRPYTEPFPGQHS 62
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 495 DIGFVIDGSSSVGTGNFRVTLQFVTMLTKFEISDTRIGAVQYTYEQRLFEFGFDKYSS 554
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 63 GEAAODAVRASAQRMDGTHGTGLALVYAKQLFAEASGARPGVPKVLWVTDGSSSDPYG 122
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 555 KPDIILAIKRVGYSQGTSGAALINPDLQELFKK---SKPKRKMILITDGKSYDVR 611
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 123 PMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLIHIVQE 174
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 612 PMAAHLKGVITTAIGVAMAAGELEVIATHPARDHSFDFEDNHLQVYPR 664
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```

RESULT 11
US-10-453-372-186
; Sequence 186, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 186
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-186
```

```

Query Match          17.9%; Score 163; DB 6; Length 709;
Best Local Similarity 31.4%; Pred. No. 1.7e-08;
Matches 54; Conservative 27; Mismatches 75; Indels 16; Gaps 5;
```

```

QY 1 RGLMLFLDSSASVSHYEFSSRVREVGQVLAPLPLGTGALRASLVHVGSRPYTEPFPGQ 60
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 66 RLFLVLVDSSSVGVNRSSELMFVKLLSDPFVPTATRAIVTFSSKNY-VERADY 124
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 61 SSCEAAQ-----DAVASAQRMDGTHGTGLALVYAKQLFAEASGARPGVPKVLWV 112
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 125 ISTRRAROKCALLLOEIPALISYRGCGTYTKGAFQAAQIIL---HARENSAKVFLIT 180
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 113 DGGSS--DVGPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFV 162
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 181 DGTSGNDP-RPLAASLRDSGVEITFTGIMQNTRELMDMASTPKKEHCYLL 231
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```

RESULT 12
US-11-080-026-4
; Sequence 4, Application US/11080026
; Publication No. US20050260192A1
; GENERAL INFORMATION:
; APPLICANT: Springer, Timothy A.
; APPLICANT: Shimooka, Motomu
; APPLICANT: Lu, Chafen
; TITLE OF INVENTION: MODIFIED POLYPEPTIDES STABILIZED IN A
; FILE REFERENCE: DESIRED CONFORMATION AND METHODS FOR PRODUCING SAME
; CURRENT APPLICATION NUMBER: US/11/080,026
; PRIOR FILING DATE: 2005-03-15
; PRIOR APPLICATION NUMBER: 09/945,265
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/229,700
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1152
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-080-026-4
```

```

Query Match          17.8%; Score 162.5; DB 7; Length 1152;
Best Local Similarity 28.1%; Pred. No. 3.6e-08;
Matches 52; Conservative 31; Mismatches 91; Indels 11; Gaps 5;

QY 3 DLMLFLDSSASVSHYEFSSRVREVGQVLAPLPLGTGALRASLVHVGSRPYTEPFPGQHS 62
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 150 DLFLDSSGSIIPDPFRMKEFVSTVWQLKSKTLF--SLMQSSEERHIFTFERON 207
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 63 GEAAODAVRASAQRMDGTHGTGLALVYAKQLFAEASGARPGVPKVLWVTDGSS-SDPYG 121
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 208 NNPNRSIVPITQLGRFTATGIRKRVRELFINTGARKAKNAFKILVITDEKFKFDPLG 267
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 122 --PMOELKDLGVTVFIVSTG-----RGNFLELSAASAPAEKHLHFV-DVDDLIHIVQE 173
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 268 YEDVLPEDADREGVIRYVIGVGDAFRSEKSRDELNTIASKPPRDHVFQVNNFALKTIQ 327
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 174 LRGS 178
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 328 LRK 332
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```

RESULT 13
US-10-453-372-180
; Sequence 180, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
```

```

1 CURRENT FILING DATE: 2003-06-03
2
3 PRIOR APPLICATION NUMBER: 09/789390
4
5 PRIOR FILING DATE: 2001-02-23
6
7 PRIOR APPLICATION NUMBER: 60/185367
8
9 PRIOR FILING DATE: 2000-03-01
10
11 PRIOR APPLICATION NUMBER: 09/823187
12
13 PRIOR FILING DATE: 2001-03-29
14
15 PRIOR APPLICATION NUMBER: 60/195792
16
17 PRIOR FILING DATE: 2000-03-10
18
19 PRIOR APPLICATION NUMBER: 09/839446
20
21 PRIOR FILING DATE: 2001-03-19
22
23 PRIOR APPLICATION NUMBER: 60/199476
24
25 PRIOR FILING DATE: 2000-03-25
26
27 PRIOR APPLICATION NUMBER: 09/863776
28
29 PRIOR FILING DATE: 2001-05-23
30
31 PRIOR APPLICATION NUMBER: 60/208263
32
33 PRIOR FILING DATE: 2000-05-31
34
35 PRIOR APPLICATION NUMBER: 09/939398
36
37 PRIOR FILING DATE: 2001-08-24
38
39 PRIOR APPLICATION NUMBER: 60/227800
40
41 PRIOR FILING DATE: 2000-08-25

```

Qy 1 RGDMFLDDSSASVSHYEFSRVRREYQGVLAPLPCTGALRASLVHVGSRPYTEFPFGQH 60
| : || : || | : : : | : : : |
Db 78 RLIELVFLVDSSSVGEVNFRSELMFVRKLSDPFVPVTATRAIVTFSSKXNVVPRVDYI 137

QY 61 SSGEAAQ-----DAVRAAQRMGDTHTGLALVYAKEOLFPAASGARPGVKULWVYTD 113
 Db 138 STRRARQHKCALLLOEIPALSYRGGTYTKGAFQOAAOILL-----HARENSTKVVFLITD 193
 QY 114 GGSS--DPVGPMPQELKDLGVTVFIVSTGRGNFELSAAASAPAEKHLHFV 162
 Db 194 GYSNGGDP-RPIAASLRDSGVEIFTFGIWQGNIRELNDMASTPKKEHCYLL 243

Search completed: February 13, 2006, 13:12:58
 Job time : 3.91304 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:44:15 ; Search time 2.24054 Seconds
(without alignments)
1468.904 Million cell updates/sec

Title: US-10-699-035A-5
Perfect score: 2380
Sequence: 1 atgcctccctgagcgcgcct.....ccgcacgcgcgtgacgcgtaa 1254

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Fgapop 10.0, Fgapext 0.5
Delop 6.0, Delext 7.0

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 194028

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-O=/abs/ABSGBER.spool/US10699035/runat_13022006_062453_25634/app_query.fasta_1
-DB=Published Applications_AA_New -QPMT=faactan -SUFFIX=rapbn -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -STAR1=1 -END=1 -MATRIX=bl0sum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abs92p
-USER=US10699035 @CGN 1 1 10 @runat_13022006_062453_25634 -NCP1=6 -ICPU=3
-NO_MMAPP -NEG_SCORES=0 -WAIT_DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA_New:*
1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pcp:*
2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pcp:*
3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pcp:*
4: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pcp:*
5: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pcp:*
6: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pcp:*
7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pcp:*
8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2114	88.8	445	6	US-10-453-372-2
2	419	17.6	3063	7	US-11-186-284-26
3	398.5	16.7	517	7	US-11-169-041-160
4	254.5	10.7	1717	7	US-11-182-016-20
5	253.5	10.7	956	6	US-10-131-826A-294
6	253.5	10.7	956	6	US-11-113-424-39
7	221.5	9.3	1366	6	US-10-821-234-1431
8	221.5	9.3	1366	7	US-11-186-284-31
9	215	9.0	1496	7	US-11-186-284-35

10	212	8.9	1464	7	US-11-000-463-243	Sequence 243, App
11	212	8.9	1464	7	US-11-186-284-28	Sequence 28, App1
12	212	8.9	1464	7	US-11-021-603-2	Sequence 2, App1
13	212	8.9	1467	6	US-10-821-234-1096	Sequence 1096, App
14	211	8.9	1464	7	US-11-000-463-243	Sequence 243, App
15	211	8.9	1464	7	US-11-186-284-28	Sequence 28, App1
16	211	8.9	1464	7	US-11-021-603-2	Sequence 2, App1
17	211	8.9	1467	6	US-10-821-234-1096	Sequence 1096, App
18	210.5	8.8	1466	6	US-11-821-234-964	Sequence 964, App
19	210.5	8.8	1466	6	US-11-186-284-33	Sequence 33, App1
20	206.5	8.7	1733	7	US-11-182-016-21	Sequence 21, App1
21	203.5	8.6	1466	7	US-11-186-284-33	Sequence 33, App1
22	202	8.5	1733	7	US-11-182-016-21	Sequence 21, App1
23	200.5	8.4	1823	6	US-10-995-561-998	Sequence 998, App
24	200.5	8.4	2102	6	US-10-995-561-990	Sequence 990, App
25	200.5	8.4	2108	6	US-10-995-561-989	Sequence 989, App
26	200.5	8.4	2157	6	US-10-995-561-991	Sequence 991, App
27	199	8.4	214	7	US-11-192-449-6	Sequence 6, App1
28	199	8.4	214	7	US-11-192-449-9	Sequence 9, App1
29	199	8.4	1717	7	US-11-182-016-20	Sequence 20, App1
30	198.5	8.3	1366	6	US-10-821-234-1431	Sequence 1431, App
31	198.5	8.3	1366	6	US-11-186-284-31	Sequence 31, App1
32	198.5	8.3	1742	7	US-11-182-016-23	Sequence 23, App1
33	197	8.3	843	7	US-11-129-104-89	Sequence 89, App1
34	196.5	8.3	3570	6	US-10-453-372-196	Sequence 196, App
35	195.5	8.2	3570	6	US-10-453-372-178	Sequence 178, App
36	195.5	8.2	3570	6	US-10-453-372-198	Sequence 198, App
37	195.5	8.2	3570	6	US-10-453-372-200	Sequence 200, App
38	195.5	8.2	3570	6	US-10-453-372-202	Sequence 202, App
39	195.5	8.2	3570	6	US-10-453-372-204	Sequence 204, App
40	195.5	8.2	3570	6	US-10-453-372-206	Sequence 206, App
41	192.5	8.1	1251	7	US-11-149-003-16	Sequence 16, App1
42	192	8.1	709	6	US-10-453-372-182	Sequence 182, App1
43	191	8.0	495	7	US-11-182-016-31	Sequence 31, App1
44	190.5	8.0	709	6	US-10-453-372-180	Sequence 180, App
45	190.5	8.0	709	6	US-10-453-372-186	Sequence 186, App

ALIGNMENTS

RESULT 1
US-10-453-372-2
; Sequence 2, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraSeqlet version 0.1

```
; SEQ ID NO 2
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-2

Alignment Scores:
Pred. No.: 2,776-113 Length: 445
Score: 2114.00 Matches: 417
Percent Similarity: 93.7% Conservative: 0
Best Local Similarity: 93.7% Mismatches: 0
Query Match: 88.8% Indels: 28
DB: 6 Gaps: 2

US-10-699-035A-5 (1-1254) x US-10-453-372-2 (1-445)

QY 1 ATGCTCCCTGGACGGCGCTGGCCCTGAGCTTGCGGCTGGCGCTGGCGGAGC 60
   |||
Db 1 MetLeuProTrpThrAlaLeuGlyLeuAlaLeuSerLeuAlaLeuAlaArgSer 20

QY 61 GCGCGGAGCGCGCTCCACACAGATCAGCCCGCGAGGAGCCTGATGTTCTGCTGAC 120
   |||
Db 21 GlyAlaGluArgGlyProProAlaSerAlaProAlaArgGlyAspLeuMetPheLeuLeuAsp 40

QY 121 AGCTCAGCCAGCGCTCTCTCACTACGAGTTCTCCCGGCTTCGAGAGTTTGAGGAGCTG 180
   |||
Db 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60

QY 181 GTGGCTCCACTGCGCTGGGAGCCGGGGCGCTGGCCAGTCTGGTGCAGCTGGGCACT 240
   |||
Db 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80

QY 241 CGGCCATACACGAGTTCCTTCGCGCAGACAGCTCGGAGTGAAGCTCCGAGATGCG 300
   |||
Db 81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 100

QY 301 GTGCGTCTTGTGCGCCAGCGCATGGGTGACACCCACACTGGCCCTGGCGCTGATGCC 360
   |||
Db 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120

QY 361 AAGGAACACTGTTGTTCTGAAGATAGAGTCCCGGCGAGGGGTGCCAAAGTGGCTGGT 420
   |||
Db 121 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuVal 140

QY 421 TGGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCGCCCGCATGAGAGCTCAAGAC 480
   |||
Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnLeuLeuLysAsp 160

QY 481 CTGGGCGTCACCGTGTTCATTGTTCAGCACCGGCGGAGGCAACTTCTGAGCTGTACGC 540
   |||
Db 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGlnLeuSerAla 180

QY 541 GTGCGCTCAGCCCGCCCGAGAAAGCACTGTTGTGAGCGTGAATACCTGCACATC 600
   |||
Db 181 AlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspAlaAspAspLeuHisIle 200

QY 601 ATTGTCCAGAGGTGAGGGGCTCATTTCTC---GCGATCGCGCGGAGAGCTCATGCC 657
   |||
Db 201 IleValGlnGlnLeuArgGlySerIleLeuAspAlaMetAlaArgProGlnGlnLeuHisAla 220

QY 658 ACGGAGATCAGCTCAGCGGCTTCGCGCTGGCTGGCCACCCCTGTCGACCGGACGCTG 717
   |||
Db 221 ThrGlnIleThrSerSerGlyPheArgLeuAlaTrpProProLeuLeuThrAlaAspSer 240

QY 718 GGTACTATGTTGCTGAGAGCTGTGGCCAGGCGCCACCGGGGCTGCAGAAGCGCCAGAG 777
   |||
Db 241 GlyTyrTyrValLeuLeuLeuValProSerAlaGlnProGlyAlaAlaArgGlnGln 260

QY 778 CTGGCCAGGGAAGCGCACGAGCTGATCTGGCGCGGCTGCAGCCGAGCACTAGCAGC 837
   |||
Db 261 LeuProGlyAsnAlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyrAsp 280

QY 838 GTGGCGCTAGTGCCTGATCCAACTGCGCGCTCTGAGGCCCGCAGATCTCGCGGCTGGC 897
```

```
Db 281 ValAlaLeuValProGlnSerAsnValArgLeuLeuAspProGlnIleLeuArgValArg 300
QY 898 ACGCGG----- 903
Db 301 ThrArgProGlyGluAlaGlyProGlyAlaSerGlyProGlnSerGlyAlaGlyProAla 320
QY 904 -----CCAGAGAGCGCGGCGCCAGACCGCATGCTCATC 936
Db 321 ProThrGlnLeuAlaAlaLeuProAlaProGlnGluAlaGlyProGlnArgIleValIle 340
QY 937 TCCGACGCGCGCGCGGAGCGCTCCGCGTGAAGTTGGGCGCCAGCGCTGGGCTCAGCGCGG 996
Db 341 SerHisAlaArgProArgSerLeuArgValSerTrpAlaProAlaLeuGlySerAlaHis 360
QY 997 GCGCTCGGCTACCAAGTTCAGCTTCGAGCGCGCTGCGCGGCGGAGCGCACCGGCTGAG 1056
Db 361 AlaLeuGlyTyrHisValGlnPheGlyProLeuMetArgGlyGlyAlaGlnArgValGln 380
QY 1057 GTGCCCGCGGCGCGCACTGCACCAAGCTGCAGGGGCTGGCGCGCGGCACTGACTG 1116
Db 381 ValProAlaGlyArgAsnGlyThrThrLeuGlnGlyLeuAlaProGlyThrAlaTyrLeu 400
QY 1117 GTGACCTGACCGCGCTTCCTCCGCTCGGCGCGCGGAGAGCGGCTGCCCAAGGCTGCG 1176
Db 401 ValThrValThrAlaAlaPheArgSerGlyArgGlnSerAlaLeuSerAlaLysAlaCys 420
QY 1177 ACGCCGAGCGGCGCGCGCGCGCGCCAGCGCCCGTGCAGCCCGCGCCGAGCGGAGC 1236
Db 421 ThrProAspGlyProArgProArgProArgProValProAlaGlnProThrProGlyThr 440
QY 1237 GCCAGCGGTGAGCGC 1251
Db 441 AlaSerArgGluPro 445

RESULT 2
US-11-186-284-26
; Sequence 26, Application US/11186284
; Publication No. US2005026493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3063
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-26

Alignment Scores:
Pred. No.: 1,5e-17 Length: 3063
Score: 419.00 Matches: 128
```

Percent Similarity:	39.9%	Conservative:	58
Best Local Similarity:	27.5%	Mismatches:	174
Query Match:	17.6%	Indels:	106
DB:	7	Gaps:	9

QY	94	CGAGGGGACCTATGTTCTGCTGGACAGCTCAGCCAGCGCTCTCTCACTACAGATTCTCC	153
Db	438	lysalaaplllevalpheuuevalapbgllyserTtySerllelylllealaapnheval	457
QY	154	CGGGTTCCGAGCTTTGTGGGGCAGCTGTGGCTCCACTGCGCCCTGGGACCGGAGCCCTG	213
Db	458	lysValAlaGAlaPheLeuGluValLeuVallySerPheGluTlleSerProAnaGVal	477
QY	214	CGTGCAGCTGTGTGACGTGGGCACTCGGCCATACAGCAATTTCCCTTGGCCAGCAC	273
Db	478	GnlleSerLeuValGlnTyrSerAArgAspProHlaSerhclupheThrleuLyLysPhe	497
QY	274	AGCTGGGGAGAGCTGCCCGAGAGCGGGTCCGCTTTCGCCAGCGCATGGGTGACACC	333
Db	498	ThrlysvaGluAspIlelleGuaIlaIleasnThrPheprolyrArglyglySerThr	517
QY	334	CACACTGCGCTGCGCTGTCTATGCCAAGAACAGCTGTTTGTGTAAGCATGAGTGCC	393
Db	518	AsnThrlyGlyAlaIleMerThrTyrValArgGluLyIlePheValProserlyGlySer	537
QY	394	CGGCCAGGGGTCCCAAAATGCTGGTGTGGGTGACAGATGGCGGCTTCCAGCACCTGTG	453
Db	538	ArgSerAsnValProLySValMetIleuIleThrAspOlyLysSerSerAAlaPhe	557
QY	454	GGCCCCCCTATCGAGCTCAAGACCTGGGGCTGACCGTTCATTGTCACACACCGC	513
Db	558	ArgAspProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValys	577
QY	514	CGAGGCAACTTCTGGAGCTGTCAAGCGCGTGCCTCAGCCCTGCGGAGAGACCTGCAC	573
Db	578	AspAlaValArgSerGluLeuGlnAlaIleAlaSerProProAlaGluThrHISValPhe	597
QY	574	TTTGTG---GAGCTGATGATCCTGCACATCATTTGTCCMAAGCTGAGGGGCTCATT---	627
Db	598	ThrValGluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIleCys	617
QY	628	-----CTCCGC-----ATGCCGCCGCGACAG	648
Db	618	LeuArgIleGluGlnGluLeuAlaAlaIleLysLybAlaIleTyrAlProProlYsAsp	637
QY	649	CTCCATGCCACGAGATCATCGTTCACGGGCTTCCGCTGAGCTGCACCCCTG----	702
Db	638	LeuSerPheSerGluValThrSerTyrGlyPheLyThrAsnThrPheProAlaGlyGlu	657
QY	703	-----CTACCCGCAACTCCGGCTACTATGTG	729
Db	658	AsnValPheSerTyrHisIleThrTyrLySglnAlaAlaGlyAspAspGluAlThrVal	677
QY	730	CTGCAG-----CTGTCGCCGACGCGCCAGCG-----	756
Db	678	ValGluProAlaSerSerThrSerValValLeuSerSerLeuLySProGluThrLeuTyr	697
QY	756	-----	756
Db	698	LeuValAsnValThrAlaGluTyrGluAspGlyPheSerIleProLeuAlaGlyGlnGlu	717
QY	757	-----GGGCTGCAAAACGCGACAGCTGCCAGAGAAAGCGCACGAGC	798
Db	718	ThrThrGluGluVallySglnAlaProArgAsnLeuLySValThrAspGluThrThrAsp	737
QY	799	-----TGATCTGGGCC-----	810
Db	738	SerPheLySleThrThrTyrGlnAlaProGlyArgValLeuArgCysArgIleIleTyr	757
QY	810	-----	810

RESULT 3
US-11-169-041-160
; Sequence 160, Application US/11169041
; Publication No. US20060019284A1

```

: GENERAL INFORMATION:
: APPLICANT: Bristol-Myers Squibb Company
: TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
: TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
: TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
: TITLE OF INVENTION: CELLS
: FILE REFERENCE: 10001 NP
: CURRENT APPLICATION NUMBER: US/11/169,041
: CURRENT FILING DATE: 2005-06-28
: PRIOR APPLICATION NUMBER: 60/584,405
: PRIOR FILING DATE: 2004-06-30
: NUMBER OF SEQ ID NOS: 527
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO: 160
: LENGTH: 517
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-11-169-041-160

```

Alignment Scores:	
Pred. No.:	2,83e-16
Score:	398.50
Percent Similarity:	45.8%
Best Local Similarity:	30.0%
Query Match:	16.7%
DB:	7
Length:	517
Matches:	114
Conservative:	60
Mismatches:	179
Indels:	27
Gaps:	6

US-10-699-035A-5 (1-1254) x US-11-169-041-160 (1-517)

Oy 106 ATGTTCCCTGCTGGACAGCTCAGCCAGCGTCTCTCTCACTACGAGTTCTCCCGGGTTCCGGAG 1655
|||:::|||||
Db 143 MetPheIeuValAspIlySerTyrIserTrieIleGlyIleAlaIaenPheValIAla 1627
|||:::|||||
Oy 166 TTGTGTGGGACAGCTGGTGGCTCCATGTGCCCTCGGGACCCGGGGCCCTGCGNCCAGCTGTG 2255
|||:::|||||
Db 163 PheIeuGlnValIeuValIlySerPheGluTleSerProIaenArgValGlnIleSerIeu 1827
|||:::|||||
Oy 226 GTGCACGTGGGACGTGGCCATACCCGAGTTCCCTTGGCGCACAGCTGGGGGTGAG 2855
|||:::|||||

```

Db      183 ValGlnTYrSerArgAspProHisThrGluPheThrLeuLeuYLeuYLeuPheThrIlyValGlu 202
Qy      286 GCTGCCAGAGATCCGATGCGTCTTCGCCAGCGCATGGTGACACCCACATCGGCGCTG 345
Db      203 AspIleIleGluAlaIleLeuThrPheProTYrArgGluYGLYSerThrsmThrGlyLys 222
Qy      346 GCGCTGCTATATCCAGAGACAGCTGTTGCTGAAGATCAAGGTCGCCGCGCCAGGGGTG 405
Db      223 AlaMetThrTYrValArgGluYLeuYLeuPheValProSerLeuYGLYSerArgSerAsnVal 242
Qy      406 CCCAAGTCTGCTGGTGGATGACAGATGGCGGCTCCAGAGCCCTGGAGCCGCCCCCGCATG 465
Db      243 ProLYValMetIleLeuIleThrAspGlyLysSerSerAspAlaPheArgAspProAla 262
Qy      466 CAGAGAGCTCAAGAGACCTGGCGCTCACCGTTCATTTGATGAGACCGCGCGCAACATTC 525
Db      263 IleuYLeuArgAsnSerAspValGluIlePheAlaValGlyValLYsAspAlaValAsp 282
Qy      526 CTGAGAGCTGTACGCCGCTGCTCAGCGCCCTCCGAGAGACCTGACATTTGTG--GAC 582
Db      283 SerGluLeuGluAlaIleAlaSerProAlaGluThrIleValPheThrValGluAsp 302
Qy      583 GTGATGACCTGCATCATTCATTCACAGAGCTGAGGGGCTCCATT----- 627
Db      303 PheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIleCysLeuArgIleGlu 322
Qy      628 -----CTCGCG-----ATCGCGCGCAGACAGCTCCATGCGCAG 660
Db      323 GlnGluLeuAlaAlaIleLYsLYsLYsAlaTYrValProProLYsAspLeuSerPheSer 342
Qy      661 GAGATCAGCTCCAGGGGCTTCGCGCTGCGCTGCGCCACCCCTGCTGACCGCAACACTGGGG 720
Db      343 GluValIThrSerTYrGlyPheLYsThrAsnTPSerProAlaGlyGluAsnValPheSer 362
Qy      721 TACTATGCTGAGAGCTGAGTCCAGAGCGCCAGCGGGCTGTCAAGACCGCAGCAGCTG 780
Db      363 TYrHisIle-----ThrTYrLYsGluAlaAlaGlyAspAspGluVal 376
Qy      781 -----CAGAGGAAACGCAACGAGCTGAGATCTGGAGCGCGCTCCACCGCGAGCAG 828
Db      377 ThrValValGluProlAserSerThrSerValValLeuAsnSerIleuLYsProGluThr 396
Qy      829 GACTACGACGTGGCGCTAGCTGAGTCAACAGTGGCGCTCTGAGGCGCCACGATCTG 888
Db      397 LeuTYrLeuValAsnValThrAlaGluTYrGluAspGlyPheSerIleProLeuAlaGly 416
Qy      889 CGGAGCGGACCGCGCAGAGAGAGCGCGCGCAGAGCGCATGTCATCCACCGCGCG 948
Db      417 GluGluThrThrGluGluValLYsGlyAlaProAlaGAsnLeuLYsValThrAspGluThr 436
Qy      949 CCGCGCAGCGCTCGCGTGAATTGGGCGCCAGCGCTGGGCTCAGCGCGCGCTCGGCTAC 1008
Db      437 ThrAspSerPheLYsIleThrSerThrGlnAlaProGlyLys-----ValLeuArgTYr 454
Qy      1009 CACGTGCACTGGCGCGCTGCGGGGCGGGGAGCGGCGCAGCGGCTGAGAGTCCCGCGGCG 1068
Db      455 ArgIleIleTYrArgProValAlaGlyGlyGluSerArgLYsValThrThrProAsn 474
Qy      1069 CGCACTGACACGAGGAGCGCTGCGCGCGCGCGGACCGCTACCTGATCGCGCTGAGC 1128
Db      475 GlnAlaGArgArgThrIleuGluAsnLeuIleProAspThrLYsTYrGluValSerValIle 494
Qy      1129 GCCGCGCTTCGCTCGGCGCGCAGAGCGGCTGTCCGCGCAAGCGCTGACGCGCGCGC 1188
Db      495 ProGluTYrPheSerGlyProGlyThrProLeuThrGlyAsnAlaAlaThrGluGluGly 514

```

```

; FILE REFERENCE: 038602/0102
; CURRENT APPLICATION NUMBER: US/11/182,016
; CURRENT FILING DATE: 2005-07-15
; PRIOR APPLICATION NUMBER: US/09/958,359
; PRIOR FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 1717
; TYPE: PR
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Grub
US-11-182-016-20

Alignment Scores:
Pred. No.: 3,29e-08 Length: 1717
Score: 254.50 Matches: 161
Percent Similarity: 31.3% Conservative: 23
Best Local Similarity: 27.4% Mismatches: 155
Query Match: 10.7% Indels: 250
DB: 7 Gaps: 39

US-10-699-035A-5 (1-1254) x US-11-182-016-20 (1-1717)
Qy      1233 CCCCGGGGTCGGGCG-----CGGGGACAGGGGCGTGGGCGCGGCGCGCGCGCTC 1183
Db      725 ProArgCysAlaGlyLeuGlyTYrAlaGlnIleGlyAlaGlyLYsSerSerArgAlaArg 744
Qy      1182 GGGCGTGACAGCGCTTGGCGGACAGCGCGCTCGCG----- 1147
Db      745 GlyArgGlyGlySerGlyAsnAlaGlnAlaThrAlaGluGlyAlaGlyArgSerProAla 764
Qy      1146 ---GCCCGAGCGGAAGC---GCCGCTCAC-----GATGAC 1117
Db      765 AspGlyThrAlaGluGlyTYrProLYsIlePheAspArgAlaAlaLeuHisSerGlnGln 784
Qy      1116 CAGTAGGCGGTCGCCGCGCGCGCGCGCGCTGACGCTGTCAGTTGCGGCGCGCGCGCGAC 1057
Db      785 GlnAlaGlyGlyProArgProSerTYrThrValSerGly-----SerGlyArgGlyHis 802
Qy      1056 CT-CCAGCGGCTGGCGCTCCCGCGCGCGCGCGCGCGCACTGACAGCTGTCAGCGCGAGCG 998
Db      803 SerProAlaCysAlaPheLeuGlnProAlaArgAla---AlaAlaArgAlaAlaAlaVal 821
Qy      997 CCGCGCTGAGCGCA-----GCCGTGGGCGCCCACTCAACGCGAGCG--- 956
Db      822 ProAlaAlaThrProAlaGlyValAlaValAlaLeuGlyPro-----ArgGlyGlyAla 839
Qy      955 -----TGGCGCGCGCGG----- 944
Db      840 AlaGlyLYsLeuCysTYrAlaIleProGlyHisLYsLeuValCysProAlaGlyAspArgAlaAla 859
Qy      943 -----CGTGGG----- 938
Db      860 IleProCysPheGlnArgGlyProGlyAlaProGlyProGlyTYrGlyProGlySer 879
Qy      937 AGATGACATGGCGCTGTG----- 920
Db      880 GlyGlyGlyCysHisLeuProGluLYsAlaGlyTYrLeuThrAlaAlaGlyAlaGlyGlu 899
Qy      919 -----GCCCGGCTTCCTGCGCGCGTGGCGCGAGCGCGAGATCTGGGCGCTCAGGA 869
Db      900 TrpProProSerGlyProAlaAlaThrAlaLeuLeuProAlaGlyTYrMetGlyGlyGly 919
Qy      868 -----GGGCGACGTTGACT-----CAG 851
Db      920 LeuCysSerAlaGlyArgSerTYrAlaGlySerGlyCysValaGlyCysThrGlyPro 939
Qy      850 GCACCTAGCGGCACGT-----CGTAGTCCGCTGCGGCT 818
Db      940 AlaAlaGlyProArgAlaGlnCysArgHisLYsLeuProGlyAspAlaGlyProGlyProGly 959

```

```

RESULT 4
US-11-182-016-20
; Sequence 20, Application US/11/182016
; Publication No. US20060019294A1
; GENERAL INFORMATION:
; APPLICANT: SUGEN, INC.
; TITLE OF INVENTION: TYROSINE KINASE SUBSTRATE (TKS) PROTEINS

```

QY 817 CGAGCGCGCCGATCCAGTCCG----- 794
Db 960 ProGlyGlnPro---SerSerProAlaArgMetGlyProLeuProGlyProLeuProArg 978
QY 794 ----- 794
Db 979 AlaArgGluGlnAspProAlaThrProGlyArgGlyGlyGlnProThrGlyLeuProThr 998
QY 793 -----TGGCGTTCC-----CTGGCA 779
Db 999 ThrAlaGlyArgArgCysGlnGlnIntPrArgGlyProValGlyValaProGlnProLeuAla 1018
QY 778 GCGTGGGCGCTTTCAGACCCCGCGCTGGCGCA-----CGAGCTCCAGCA 728
Db 1019 GlnProGlnLeuLeuAlaAlaProGlnGlnProThrAlaThrAlaSerProIleProLeu 1038
QY 727 CATAGAGCCCGAGTGGCGGTGAGGAGGTGGCCAGGCGGAGCGCGTGGAGG 668
Db 1039 LeuProGlyProMetTrpArgGlyLeuGly----- 1048
QY 667 TGATCTCCGTGGATGAGCTGGCGCGCATCCGAGMATGAGCCCTCAGCTCTT 608
Db 1049 -----ArgGlyProAlaGlySerArgSerArgGlyGlnAlaProLysSer--- 1063
QY 607 GGAATATGATGTCAGAGTATCCAGTCCAGAAAGTCAGAGTCTTCTCGGCGAGGCGTG 548
Db 1064 -----CysAla---AspProArgPro-----GlyGlyHisGlnHisGlyGly 1076
QY 547 AGGCAAGGCGTGCACGCTCCAGGAAGTTGCTGGCGCGGTGCAGCAATGAACAGGTGA 488
Db 1077 ArgGlnAspValLeuThrThrGly-----ThrArg--- 1086
QY 487 CGCCGAGCTCTTGAAGCTCTGCATGGGAGGCG----- 455
Db 1087 AlaAlaGlyPro-----GlyGlyAlaArgArgThrLeuGlySerArgArg 1101
QY 454 ---CCAGAGGCTGCTGGAGCCGCCATCTGTACCCAGCACAGCACTTGGGCAACC--- 401
Db 1102 ProProAspGlyValaGlnAlaGlyHisGlnCysProAlaAlaAlaGlyValaAlaAspCys 1121
QY 400 ---CTGGCGGCGACCTGATGCTTACAGCAAAACAGCTTCTGGCGCATAGACCA 350
Db 1122 LeuThrArgLeuArgGlyHis-----LeuGlnAlaSerAlaThrProTrpAlaAlaAsp 1139
QY 349 GCG---CCAGGCGAGGAGTGGGTGTACCCA---TGGCGTGGCGAGAGACAGCAGCGCAT 296
Db 1140 AlaThrSerGlyHisLeuGlyCysCysProGlnCysProGly---LysAlaSerGlnLeu 1158
QY 295 CCGTGGGAGCTTACCCGAGAGCTGTGGCGGAGGAGAACTCGGTGTATGGCGGACTGC 236
Db 1159 Pro-----ProAspThrLeuSerAlaGlyAlaSerGlyLeuArg-----His 1172
QY 235 CCAAGTGCACAGACAGTGGACGAGAGGCGCGCGTGGCGGCGGCGAGTGGAGCCA----- 182
Db 1173 ProProProThrHisTrpGlyLeuLeuProSerProArgGlyProValGlnProLeuCys 1192
QY 181 -----CCAGCTGCCCAACAATCCGGAACCCGGAGAGACTGTAGTAGAGACGC 131
Db 1193 ThrValArgGlnAlaProThrGlnThrGlyGlu----- 1203
QY 130 TGGCTAGAGCTGTCAGACAGGA---ACATCAGTCCCTCGGGGGGCGTATGCTGTGGAC 74
Db 1204 TrpSerGlyCysAlaGlnSerLeuLysGlnGlyLeuHisGlyLysTrpProLeu----- 1221
QY 73 CGGCTCCGCGCGCTCCGCGCAGGCGCGAGCGCGCAAGCTCAGAGG----- 29
Db 1222 -----ProAlaProSerProAlaAlaAlaSerGlyThrAlaAspSerVal 1236
QY 28 -----CCAGGCGGAGCGCG 14
Db 1237 TrpAlaAlaProGlyGlyAlaPro 1244

US-10-131-826A-294
; Sequence 294, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RJC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 294
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-294
Alignment Scores:
Pred. No.: 4.11e-08 Length: 915
Score: 253.50 Matches: 62
Percent Similarity: 50.0% Conservative: 36
Best Local Similarity: 31.6% Mismatches: 85
Query Match: 10.7% Indels: 13
DB: 6 Gaps: 3
US-10-699-035A-5 (1-1254) x US-10-131-826A-294 (1-915)
QY 46 GGGCTGGCGGAGAGCGCGCGGCTCCAGCAGATCAGCCCCCGAGGGGAGCCTG 105
Db 45 AlaLeuLeuGlnSerSerCysGlnuSn-----LysArgAlaAspLeu 58
QY 106 ATGTTCTGCTGAGACAGCTCAGCAGCGGTCTCTCACTAGAGTTCTCCGGGTTGGAG 165
Db 59 ValPheIleIleAspSerSerArgSerValaSnThrHisAspTyrAlaLysValLysGlu 78
QY 166 TTGTGGGCGAGCTGTGTGCTCACTGCGCCCTGGGCGACCGGGGCCCTGCGTGCAGTCTG 225

```

Db      79 PheIIeValAspIIeLeuGlnPheLeuAspIIeGlyProAspValThrArgValGlyLeu 98
Qy      226 GGCACGTCGGCAGTCGGCCATACACCGAGTTCCTCCCTTCGGCAGACAGCTCGGGTAG 285
Db      99 LeuGlnTYrGlySerThrValIysAsnGluPheSerLeuIleThrPheIysArgIysSer 118
Qy      286 GCTGCCAGAGATCGCGTGCCTCTCTGCCAGCGCATGGTGACACCCACATCGGCTG 345
Db      119 GluValAGluArgAlaValIysArgMetArgHisLeuSerThrGlyThrmethrGlyLeu 138
Qy      346 GGGCTGCTGTATGCCAAGAACAGCTGTTGCTGAAGCATCAGTCCCGGCCA----- 399
Db      139 AlaIleGlnTYrAlaLeuAsnIleAlaPheSerGluAlaGluGlyAlaArgProLeuArg 158
Qy      400 ---GGGGTCCCAAGTCTGGTGTGGTGACAGATGGCGGCTCCAGCAGCCTGTGGGC 456
Db      159 GluAsnValProArgValIleMetIleValThrAspGlyArgProGlnAspSerValAla 178
Qy      457 CCCCCCATGACAGAGCTCAAGACCTGGGCGTCCACCGTTCATTTGTGACACCGCCGA 516
Db      179 GluValAlaAlaIylsAlaArgAspThrGlyIleLeuIlePheAlaIleGlyValGlyGln 198
Qy      517 GGCACCTTCCTGAGAGCTGTACCGCTGCTCAGCCCTCGCGAGAGACCTG----- 570
Db      199 ValAspPheAsnThrLeuIysSerIleGlySerGluProHisGluAspHisValPheLeu 218
Qy      571 -----CACTTGTGACGTGATGACCTGCACATCATTTGTCACAG 612
Db      219 ValAlaAsnPheSerGlnIleGlnThrLeuThrSerValPheGlnIys 234

```

RESULT 6
US-11-113-424-39

```

; Sequence 39, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangoli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acid Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; PRIOR FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 39
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-39

```

Alignment Scores:
Pred. No.: 4.09e-08 Length: 956
Score: 253.50 Matches: 62
Percent Similarity: 50.0% Conservative: 36
Best Local Similarity: 31.6% Mismatches: 85
Query Match: 10.7% Indels: 13

DB: 7 Gaps: 3

US-10-699-035A-5 (1-1254) x US-11-113-424-39 (1-956)

```

Qy      46 GCGTGGGGGAGGAGGGGGGGGGGCGGCTCCACAGATCAGCCCCCGAGGGAGCTG 105
Db      45 AlaLeuLeuGluSerSerCysGlnAsn-----LysArgAlaAspLeu 58
Qy      106 ATGTTCTGCTGAGACAGCTCAGCCAGCGTCTCTCACTACAGATTCTCCGGGTTGGAG 165
Db      59 ValPheIleIleAspSerSerArgSerValAsnThrHisAspTYrAlaIylsValIysGln 78
Qy      166 TTGTGGGGGACAGCTGTGCTTCATGCTGCCCTCGGACCGGGGCGCTCGTCCAGTCTG 225
Db      79 PheIIeValAspIIeLeuGlnPheLeuAspIIeGlyProAspValThrArgValGlyLeu 98
Qy      226 GGCACGTCGGCAGTCGGCCATACACCGAGTTCCTCCCTTCGGCAGACAGCTCGGGTAG 285
Db      99 LeuGlnTYrGlySerThrValIysAsnGluPheSerLeuIleThrPheIysArgIysSer 118
Qy      286 GCTGCCAGAGATCGCGTGCCTCTCTGCCAGCGCATGGTGACACCCACATCGGCTG 345
Db      119 GluValAGluArgAlaValIysArgMetArgHisLeuSerThrGlyThrmethrGlyLeu 138
Qy      346 GCGCTGCTGTATGCCAAGAACAGCTGTTGCTGAAGCATCAGTCCCGGCCA----- 399
Db      139 AlaIleGlnTYrAlaLeuAsnIleAlaPheSerGluAlaGluGlyAlaArgProLeuArg 158
Qy      400 ---GGGGTCCCAAGTCTGGTGTGGTGACAGATGGCGGCTCCAGCAGCCTGTGGGC 456
Db      159 GluAsnValProArgValIleMetIleValThrAspGlyArgProGlnAspSerValAla 178
Qy      457 CCCCCCATGACAGAGCTCAAGACCTGGGCGTCCACCGTTCATTTGTGACACCGCCGA 516
Db      179 GluValAlaAlaIylsAlaArgAspThrGlyIleLeuIlePheAlaIleGlyValGlyGln 198
Qy      517 GGCACCTTCCTGAGAGCTGTACCGCTGCTCAGCCCTCGCGAGAGACCTG----- 570
Db      199 ValAspPheAsnThrLeuIysSerIleGlySerGluProHisGluAspHisValPheLeu 218
Qy      571 -----CACTTGTGACGTGATGACCTGCACATCATTTGTCACAG 612
Db      219 ValAlaAsnPheSerGlnIleGlnThrLeuThrSerValPheGlnIys 234

```

RESULT 7

```

US-10-821-234-1431
; Sequence 1431, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: PC_SEQ_genes Version 1.0
; SEQ ID NO 1431
; LENGTH: 1366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1431

```

Alignment Scores:
Pred. No.: 2.5e-06 Length: 1366
Score: 221.50 Matches: 135
Percent Similarity: 34.8% Conservative: 28
Best Local Similarity: 28.8% Mismatches: 173
Query Match: 9.3% Indels: 133

Db 789 GlyAsnAspGlyAlaGlyLeuProGlyProLeuGlyProProGlyProAlaGlyLeu 808
QY 123 CTCAGCAGCGTCTCTCTACCTACGAGTTCTCCGGGT---TCGGAGTTTGGGGAGCT 179
Db 809 LeuGlyGlyLeuGlyGlyLeu-----ProGlyProAlaGlyLeuValGlyProPro 824
QY 180 GGTGGCTCAGCTGCGCTTGGGACCGGGGGCTT----- 212
Db 825 GlySerArgGlyAsnProGlySerArgGlyGluAsnGlyProThrGlyAlaValGlyPhe 844
QY 213 GGTGGCAGCTGAGTGGGACGAGTGGGACGACGAGTTCCCTTGGGGCAGCA 272
Db 845 AlaGlyProGlnGlySerAspGlyGlnProGlyValGlyGlyGlnProGlyGlnProGly 864
QY 273 CAGCTCGG---TGAGCTGCCAGGA---TGCAGTGCCTTCTTCCAGCGCATGGG 326
Db 865 GlnIlySerGlyAspAlaGlySerProGlyProGlnGlyLeuAlaGlySerProGly----- 882
QY 327 TGACACCACTGAGCTGGGCGCT-----GGTCTATGCCAGGAACAGCTGTTGCTGA 380
Db 883 -----ProHisGlyProAsnGlyValProGlyLeuIlyGly----- 895
QY 381 AGCATAGTGGCCCCGGGCGGTCGCCAAAGTCTGTGTGGTACAGATGGCGCTC 440
Db 896 -----ArgIlyThrGln----- 899
QY 441 CAGCAGCCTGTGGGCCCCCATGCA-----GAGCTCAAGA----- 479
Db 900 -----GlyProProGlyAlaThrGlyPheProGlySerAlaGlyValGly 915
QY 480 ---CCTGGGCGTCAACCGGTTCATTGTACAGCCGCGAGGCACTTCTGGAGCTGTC 536
Db 916 ProGly-----ProAlaGlyAlaProGlyProAla 926
QY 537 AGCCGTGCTCAGCCCTGCGAGAGACCTGCACTTGTGAGATGATGACTGCA 596
Db 927 GlyProLeuGlyGlnProGlyLysGlnGlyPro-----ProGly 939
QY 597 CATCATTTTCAGAGCTGAG---GGGCTCATTTCTCGGATGCGCGCAGCAGCTCA 653
Db 940 -----ProArgIlyAspProGlySerHisGlyArg---ValGlyValaArgGlyPro 955
QY 654 TCCCAAGAGATCAGTCCAGCGGCTTCCGCTGCGCTGCGCACCCTGCTGACGCA 713
Db 956 AlaGly-----ProGly-----GlyProGlyAspIlyGly 966
QY 714 CTCGGGCTACTATGCTGTGAGCTGTGCTCCAGCGCCAGCGGGGCTGCAAGACCA 773
Db 967 -----AspProGlyGluAspGlyGlnProGlyProAspGlyProProGlyPro 982
QY 774 GAGCTGCGAGGAGACGCGAGCTGATCTGGGCGGCGCTGACCGGAGACAGGACTA 833
Db 983 AlaGlyThrGlyGlnArgIlyIle---ValGlyMetProGly----- 996
QY 834 CAGCTGGGCTAGTGCCTGAGTCCAGCTGCGCTCTGAGGCGCCAGATCTCTGGGCT 893
Db 997 GlnArgGly-----GluArgGlyMetProGlyLeuProGlyProAlaGly 1011
QY 894 GCGCAGCGCGCAGAGAGCGCGGCGCAGAGCATGTCATCTCCACGCGCGCGCG 953
Db 1012 ThrProGlyLysValGlyProThrGlyAlaThrGlyAspIlyGlyProProGlyProVal 1031
QY 954 CAGCCTCGCGTGAAGTTGGGCGCCAGCGCTGGGCTGACGCGCGCTCGGCTACACGT 1013
Db 1032 GlyProProGlySerAsnGlyProValGly-----GluProGlyProGlnGlyProAla 1049
QY 1014 GAGATTGCGCGCTGCGCGCGGAGCGCAGCGAGT---GAGAGTCCCGCGG--- 1067
Db 1050 GlyAsnAspGlyThrProGlyArgAspGlyAlaValGlyGlnArgGlyAspArgGlyAsp 1069
QY 1068 -----CCGCACTGCACACGCTGCAGG----- 1091
Db 1070 ProGlyProAlaGlyLeuProGlySerGlnGlyAlaProGlyThrProGlyProValGly 1089

QY 1092 ---CCTGCG---GCCGAGCAGCGCTTACCTGTGACCGGTGACCGCGCTTCCGCTGCG 1145
Db 1090 AlaProGlyAspAlaGlyGlnArg-----GlyAspProGlySerArgGlyProIlyGly 1107
QY 1146 CCGCAGAGCGCGCTGTCTCCGCCAAGGCTTCACGCGCCGACGCGCGCGCGCCAG 1205
Db 1108 HisLeuGlyArgAlaGlyLysArgGlyLeu-----ProGly 1119
QY 1206 CCGGTGCGCGCGCGCGCGCGCGCGAGC 1235
Db 1120 ProGlnGlyProArgGlyAspIlyGlyAsp 1129

RESULT 10

US-11-000-463-243
; Sequence 243, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhimei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIPACN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-000-463-243

Alignment Scores:

Pred. No.: 8,52e-06 Length: 1464
Score: 212.00 Matches: 140
Percent Similarity: 33.6% Conservative: 31
Best Local Similarity: 27.5% Mismatches: 168
Query Match: 8.9% Indels: 170
DB: 7 Gaps: 33

US-10-699-035a-5 (1-1254) x US-11-000-463-243 (1-1464)

QY 15 GCGCTCGG-----CCTGCGCTGAGCTTCCGCTGCG--- 47
Db 572 GlyAlaArgGlyGlnAlaGlyValMetGlyPheProGlyProIlyGlyAlaIlyGln 591
QY 48 ---GCTGGCGGAGAGCGCGCGGAGCGGCTCACC-----AGCATCAGCCCC----- 92
Db 592 ProGlyLysAlaGlyGlnArgGlyValProGlyProProGlyAlaValGlyProAlaGly 611

```
QY 93 -----CCGAGGGACCTGATGTTCTCTGAGACGCTC 125
Db 612 LysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAlaGlyProAlaGlyGluArg 631
QY 126 AGCCAGCGCTCTCACTACGAGTTCCGGGGTTGGGGGCTTTGGGGGCACTGCTGGCC 185
Db 632 GlyGluGln-----GlyProAlaGlySerProGlyPhe 642
QY 186 TCCACTGCCCCCTGGGACCGGGGCCCTGGCTGCATCTGTGTCAGCGTGGCACTGGCC 245
Db 643 GlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysProGlyGlnGlnGly 662
QY 246 ATAACACGAGTTCCCTTGGCCAGACAGCTCGGTGGCTGCCCAAGATGCGGTGGC 305
Db 663 Val-----ProGlyAspLeu-----GlyAlaProGlyProSerGly 674
QY 306 TGCTTCTGCCAGCGGATGGTGAGACCCACACTGGCTGGGGCTGTGTATGGCAAGA 365
Db 675 -----AlaArgGly--GluArgGlyPheProGlyGluArgGlyValGlnGly 689
QY 366 ACAGCTGTTTGTGAAGCATCAGATGTCGCCGCCGCC-----AGGGGTGCCCAAGTGTGCT 419
Db 690 ProPro-----GlyProAlaGlyProArgGlyAlaAlaGlnGlyAlaPro 703
QY 420 GTGGGTACAGATGGCGGCTCCAGCGACCTGTGGGCCGCCCAATGCA--GAGACTCA 476
Db 704 GlyAsnAsp-----GlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGln 721
QY 477 GGA-----CCTGGGGCGTACCGGTGTCTATTGTACGACCGGCCGAGCAATTCCTGTGAGCT 533
Db 722 GlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyPro 741
QY 534 GTCAAGCCGCTGCTCAGCCCTGCGCAGAAACCTGCACTTTGTGAGCTGATGATGACT 593
Db 742 -----LysGlyAspArg 745
QY 594 GCACATCATTTGTCAGA--GCTGAGGGGCTTCATTCTCGGATGCGGCC----- 641
Db 746 GlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGlyValAlaArgGlyLeu 765
QY 642 -----GCAGACGCTCCATGCGAGGATACAGCTCGAG 674
Db 766 ThrGlyProIleGlyProProGlyProAlaGlyAlaPro-----GlyAspLysGlyGlu 783
QY 675 CGGCTTCCGCTGGCTGGCCACCCCTGCTGACCGCAGACTCGGGCTACTATGTGTGGA 734
Db 784 SerGlyProSerGlyProAlaGlyProThrGlyAlaArg-----Gly 797
QY 735 GCTGGTGGCCAGCGCCAGCGGGGGCTGCAGACCGCAGCTGCGCAGGAAAGCCAC 794
Db 798 AlaProGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGlyProPro 817
QY 795 GGACTGGATCTGGGCGCGGCT-----CGACCGGACACGCACTACAGCTGGC 842
Db 818 GlyAlaAsp--GlyGlnProGlyAlaLysGlyLysProGlyAspAlaGlyAlaLysGly 836
QY 843 GCTAGTCTGAGTCCACAGCTGGGCTCTGAGGCCCGCCAGACTCTGCG----- 890
Db 837 -----AspAlaGlyProProGlyProAlaGlyProAlaGlyProProGly 851
QY 891 -----GTGCGCACCGCGCCAGAGAGAGCGCG----- 917
Db 852 ProIleGlyAsnValGlyAlaProGlyAlaLysGlyAlaArgGlySerAlaGlyProPro 871
QY 918 -----GCCAGAGCGCATCGTCACTCTCCAGCCCGCGC----- 950
Db 872 GlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsn 891
QY 951 GCGAGAGCTTCGCGGTAGTTGGGCCCGCAGCGCTGAGCTGAGCGCGCGCTGACCA 1010
Db 892 AlaGlyProProGlyProProGlyProAlaGly-----LysGlnGlyGlyLysGlyPro 909
QY 1011 CGTGACGTTCCGGGCGCTGCGGGGCGGGAGCGCAGCGGGT----- 1052
```

```
Db 910 ArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProProGlyProProGly 929
QY 1053 -----GAGGTGCGCGCGCGCCGCAACTGACACCACTGCTGAGG 1091
Db 930 ProAlaGlyGluLysGlySerProGlyAlaAspGlyPro-----AlaGlyAla 945
QY 1092 CCTGCG-----GCCGGACCGCG-----CTACTGTGT 1118
Db 946 ProGlyThrProGlyProGlnGlyIleAlaGlyGlnArgGlyValGlyValGlyLeuProGly 965
QY 1119 GACCTGACCGCGCG-----CTTCGCTCGGGCGCGCGAGACCGCGCTCCGC 1166
Db 966 GlnArgGlyGluArgGlyPheProGlyLeuPro--GlyProSerGlyGluProGlyLys 984
QY 1167 CAAGCGCTGACAGCGCCGCGCGCGCCAGCGCCCGCGCGCGCGCGAG 1226
Db 985 GlnGly-----ProSerGlyAlaSerGlyGluArgGlyPro--ProGlyP 999
QY 1227 CCGGGGACCGCGCAGCGCTGAGC 1249
Db 999 romerGlyProProGlyLeuAla 1006

RESULT 11
US-11-186-284-28
; Sequence 28, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-28

Alignment Scores:
Pred. No.: 8,52e-06 Length: 1464
Score: 212.00 Matches: 140
Percent Similarity: 33.6% Conservative: 31
Best Local Similarity: 27.5% Mismatches: 168
Query Match: 8.9% Indels: 170
DB: 7 Gaps: 33

US-10-699-035A-5 (1-1254) x US-11-186-284-28 (1-1464)
QY 15 GCGGCTCGG-----CCTGACCTGAGCTTGGCGGTGCGC--- 47
Db 572 GlyAlaAspGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyGlu 591
QY 48 ---GCTGGCGCGAGCGCGCGCGAGCGCGGTCCACC-----AGCATGACCCC----- 92
```

```

Db      522 ProGlyLyAspAlaGlyGluArgGlyValProGlyProProGlyAlaValGlyProAlaGly 611
Qy      93 -----CCGAGGGGACCTGATGTTCTCTCTGAGACGTC 125
Db      612 LysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAlaGlyProAlaGlyGluArg 631
Qy      126 AGCCAGCGCTCTCACTACAGAGTTCTCCGGGTTCCGGAGTTGNGGGGACGCTGTGCG 185
Db      632 GlyGluGln-----GlyProAlaGlySerProGlyPhe 642
Qy      186 TTCACCTGCCCTTGGGACCGGGGCTTGGCTGCACTGTGTGACAGTGTGCGACGTCGAGTCGCGC 245
Db      643 GlnGlyLeuProGlyProAlaGlyProGlyGluAlaGlyLysProGlyGlnGlnGly 662
Qy      246 ATACACCGAGTTCCCTTCCGCGACAGACGCTGCGGTGAGCGGCGCCAGAGTGGCGGTGCG 305
Db      663 Val-----ProGlyAspLeu-----GlyAlaProGlyProSerGly 674
Qy      306 TGCTTCTGCGCAGCGCATGAGTGACACCGACACTGCGCTGCGCTGTATGCCAAGA 365
Db      675 -----AlaArgGly-----GluArgGlyPheProGlyGluArgGlyValGlnGly 689
Qy      366 AAGCTGTTTGTGTAAGCATCAGAGTCCCGCGC-----AGGGGTGCCCAAGCTGTGT 419
Db      690 ProPro-----GlyProAlaGlyProArgGlyAlaAsnGlyAlaPro 703
Qy      420 GTGGGAGACAGATGGCGGCTCCAGACCGCTGTGGGCGCCCGCCATGCA---GGAGCTCAA 476
Db      704 GlyAsnAsp-----GlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGln 721
Qy      477 GGA---CCTGGGCGTACCGGTGTTGTCAGACACCGGCGGCACTTCTCTGAGCT 533
Db      722 GlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyPro 741
Qy      534 GTACGCGCTGCTCAGCCCTGCGGAGAACACTGTCACTTGTGAGACTGATACCT 593
Db      742 -----LysGlyAspArg 745
Qy      594 GCACATCATTTGCCAAGA---GCTGAGGGGCTCCATTTCGGAGTGGCGC----- 641
Db      746 GlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGlyValArgGlyLeu 765
Qy      642 -----GCAGCAGCTCCATGCGACCGAGATCAGCTCCAG 674
Db      766 ThrGlyProLleGlyProProGlyProAlaGlyAlaPro-----GlyAspLysGlyGlu 783
Qy      675 CGGCTTCCGCTGGCGCTGCGCACCCCTGCTGACCGGACCTGGGCTACTATGTCTGA 734
Db      784 SerGlyProSerGlyProAlaGlyProThrGlyAlaArg-----Gly 797
Qy      735 GGTGTGCCACGACCGCCGCGGCTGCGACAGCGCAGCTGCCAGGAAGCCAC 794
Db      798 AlaProGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGlyProPro 817
Qy      795 GAGACTGATCTGGGCGCGCT-----CGACCGGACACCGACTACGACGCTGCG 842
Db      818 GlyAlaAsp---GlyGlnProGlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGly 836
Qy      843 GCTAGTGCCTGAGTCCACAGCTGCGCTCTGAGGCGCCAGATCTGCG----- 890
Db      837 -----AspAlaGlyProProGlyProAlaGlyProAlaGlyProProGly 851
Qy      891 -----GGTGGCGACGCGGCGGACGAGGAGCGCG----- 917
Db      852 ProLleGlyAsnValGlyAlaProGlyAlaLysGlyAlaArgGlySerAlaGlyProPro 871
Qy      918 -----GCCAGAGCGCATCTCATCTCCACGCGCGCGC----- 950
Db      872 GlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsn 891
Qy      951 GCGCAGCTCCCGGTGAGTTGGGCGCCAGCGCTGGGCTAGCCGCGCTCGGCTAACCA 1010

```

```

Db      892 AlaGlyProProGlyProProGlyProAlaGly-----LysGlyGlyGlyLysGlyPro 909
Qy      1011 CDTGCAATTCGGGCGCGCTCCGCGGCGGAGCGGCGAGT----- 1052
Db      910 ArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProProGlyProProGly 929
Qy      1053 -----GAGGTGCGCCGCGGCGCAACTGACCACTGACGAG 1091
Db      930 ProAlaGlyGluLysGlySerProGlyAlaAspGlyPro-----AlaGlyAla 945
Qy      1092 CCTGCG-----GCCGGGACCGCG-----CTACCTGTGT 1118
Db      946 ProGlyThrProGlyProGlnGlyAlaAlaGlyGlnArgGlyValAlaGlyLeuProGly 965
Qy      1119 GACCGTGACCGCGCG-----CTTCCGCTCGGCGCGGAGCGCGCTGTCCGC 1166
Db      966 GlnArgGlyGluArgGlyPheProGlyLeuPro-----GlyProSerGlyGluProGlyLys 984
Qy      1167 CAAGCTGTGACGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1226
Db      985 GlnGly-----ProSerGlyAlaSerGlyGluArgGlyPro--ProGlyP 999
Qy      1227 CCGGGGACCGCGCAGCGCGGAGC 1249
Db      999 romecGlyProProGlyLeuAla 1006

RESULT 12
US-11-021-603-2
; Sequence 2, Application us/11021603
; Publication No. US20060003954A1
; GENERAL INFORMATION:
; APPLICANT: Berl, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/11/021, 603
; CURRENT FILING DATE: 2004-12-21
; PRIOR APPLICATION NUMBER: US/10/149, 352
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 4.0
; SEQ ID NO 2
; LENGTH: 1464
; TYPE: PRF
; ORGANISM: Homo sapiens
US-11-021-603-2

Alignment Scores:
Pred. No.: 8, 52e-06 Length: 1464
Score: 212.00 Matches: 140
Percent Similarity: 33.6% Conservative: 31
Best Local Similarity: 27.5% Mismatches: 168
Query Match: 8.9% Indels: 170
DB: 7 Gaps: 33

US-10-699-035A-5 (1-1254) x US-11-021-603-2 (1-1464)
Qy      15 GCGCGCTCGG-----CCTGGCGCTGAGCTTCCGCGTGC--- 47
Db      572 GlyAlaArgGlyGluAlaGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyGlu 591
Qy      48 ---GCTGGCGGAGCGGCGGAGGCGCGGTCCAC---AGCATGAGCCCC----- 92
Db      592 ProGlyLysAlaGlyGluArgGlyValProGlyProProGlyAlaValGlyProAlaGly 611
Qy      93 -----CCGAGGGGACCTGATGTTCTCTGAGACGTC 125
Db      612 LysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAlaGlyProAlaGlyGluArg 631
Qy      126 AGCCAGCGTCTCACTACAGAGTTCTCCGGGTTGNGGGGACGCTGTGCG 185

```

```

Db      632 GlyIuGln-----GlyProAlaGlySerProGlyPhe 642
Qy      166 TCACATGCGCTGGGACCGGGCCCTGCGTCCAGTGTGACAGTGGCGGCGC 245
Db      643 GInGlyLeuProGlyProAlaGlyProProGlyGlyAlaGlyLeuProGlyGlnGly 662
Qy      246 ATACACCGAGTTCCCTTGGCCAGACAGCTGGGTGAGCGTCCAGAGATGCGGTGCG 305
Db      663 Val-----ProGlyAspLeu-----GlyAlaProGlyProSerGly 674
Qy      306 TGCTTCTGCGCAGCGCATGGGTGACCCACACTGCGCTGGCGTGTATGACCAAGA 365
Db      675 -----AlaArgGly-----GluArgGlyPheProGlyGluArgGlyValGlnGly 689
Qy      366 ACAGCTGTTGTGTAAGCATCAGTGTGCGCGCC-----AGGGGTGCGCCAAAGTGTGCT 419
Db      690 ProPro-----GlyProAlaGlyProAlaGlyAlaAlaGlnGlyAlaPro 703
Qy      420 GTGGGTGACAGATGCGCGCTCCAGCGACCTGTGGGCCCCCAGCA---GAGAGCTCAA 476
Db      704 GlyAsnAsp-----GlyAlaGlySerGlyAspAlaGlyAlaProGlyAlaProGlySerGln 721
Qy      477 GGA---CCTGGGCGTACCGGTGTTCATTGTGACAGCCGGCCAGGAACCTTCTGTGAGCT 533
Db      722 GlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyPro 741
Qy      534 GTACAGCGGTGCTGACCCCTGCGCAGAAAGACCTGACTTGTGAGTGAGTACCT 593
Db      742 -----LysGlyAspArg 745
Qy      594 GCACATCATTTGCCAAGA---GCTGAGGGGCTCCATTGCGGATGCGGCGC----- 641
Db      746 GlyAspAlaGlyProGlySerGlyAlaAspGlySerProGlyAlaAspAlaGlyAlaArgGlyLeu 765
Qy      642 -----GAGACGCTTCATGCCACGGAATACAGTCCAG 674
Db      766 ThrGlyProIleGlyProProGlyProAlaGlyAlaPro-----GlyAspLysGlyGlu 783
Qy      675 CGGCTTCCGCTGGCTGGCCACCCCTGCTGACCGGAGCTGCGGCTACTATGTGCTGGA 734
Db      784 SerGlyProSerGlyProAlaGlyProThrGlyAlaArg-----Gly 797
Qy      735 GCTGTGCTCCAGCGCCAGCGCGGGGCTGCAGAGCCGACGAGCTGCCAGGAAGCCGAC 794
Db      798 AlaProGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGlyProPro 817
Qy      795 GGAAGTGAATCTGGGCGCGCT-----CGACCGGAGCAGCGACTACGACGTGGC 842
Db      818 GlyAlaAsp---GlyGlnProGlyAlaLysGlyGluProGlyAlaAspAlaGlyAlaLysGly 836
Qy      843 GCTAGTGCCTGAGTCCAGCGTGGCTCTGAGAGCCCGACAGATCCTGCG----- 890
Db      837 -----AspAlaGlyProProGlyProAlaGlyProAlaGlyProProGly 851
Qy      891 -----GTCGCGCAGCGCGCCAGAGAGCGCG----- 917
Db      852 ProIleGlyAsnValGlyAlaProGlyAlaLysGlyAlaArgGlySerAlaGlyProPro 871
Qy      918 -----GCCAGAGCGCATCTGCTATCTCCACAGCCCGCGC----- 950
Db      872 GlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsn 891
Qy      951 GCGCAGCTCCGCGATGTTGGGCCCCAGGCGTGGGCTGAGCGCGCGCTGCGATACCA 1010
Db      892 AlaGlyProProGlyProProGlyProAlaGly-----LysGlnGlyGlyLysGlyPro 909
Qy      1011 CGTGCAATTGCGGCGCTGCGCGGCGGAGCGCGACGCGGT----- 1052
Db      910 ArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProProGlyProProGly 929
Qy      1053 -----GAGGTGCGCGCGCGCGCGCGCACTGACACCACTGTGAGGG 1091

```

```

Db      930 ProAlaGlyGlyLysGlySerProGlyAlaAspGlyPro-----AlaGlyAla 945
Qy      1092 CCTGCG-----GCCGGGACCGC-----CTACCTGCT 1118
Db      946 ProGlyThrProGlyProGlnGlyIleAlaGlyGlnArgGlyValAlaGlyLeuProGly 965
Qy      1119 GACCGTGACCGCGC-----CTTCCGCTCGGGCGCGCGAGAGCGGCTGTCCGC 1166
Db      966 GlnArgGlyGluArgGlyPheProGlyLeuPro---GlyProSerGlyGluProGlyLys 984
Qy      1167 CAAGCGCTGACAGCGCCGCGCGCGCGCGCCAGCGCCCGCGCGCGCGCGC 1226
Db      985 GlnGly-----ProSerGlyAlaSerGlyGluArgGlyPro---ProGlyP 999
Qy      1227 CCGGGGACCGCGCAGCGCTGAGC 1249
Db      999 rometGlyProProGlyLeuAla 1006

RESULT 13
US-10-821-234-1096
; Sequence 1096, Application US/10821224
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmant, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIORITY FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1096
; LENGTH: 1467
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1096

Alignment Scores:
Pred. No.: 8,52e-06 Length: 1467
Score: 212.00 Matches: 140
Percent Similarity: 33.6% Conservative: 31
Best Local Similarity: 27.5% Mismatches: 168
Query Match: 8.9% Indels: 170
Gaps: 33
DB: 6

US-10-699-035A-5 (1-1254) x US-10-821-234-1096 (1-1467)
Qy      15 GCGGCTCG-----CCTGGCCTGAGCTTGGCGTGGC--- 47
Db      575 GlyAlaArgGlyGlyAlaGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyGlu 594
Qy      48 ---GCTGCGCGGAGCGCGCGGAGCGCGCTCAC---AGCATGACGCC----- 92
Db      595 ProGlyLysAlaGlyGluArgGlyValProGlyProProGlyAlaValGlyProAlaGly 614
Qy      93 -----CCGAGGGAGCATGATGTTCTGCGTGGAGAGCTC 125
Db      615 LysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAlaGlyProAlaGlyGluArg 634
Qy      126 AGCCAGCGCTCTCACTACGAGTCTCCGCGGTTGCGGAGTTTGGGGCAGCTGGTGGC 185
Db      635 GlyIuGln-----GlyProAlaGlySerProGlyPhe 645
Qy      186 TCACATGCGCTGGGACCGGGGCTTGGTGCAGTGTGACAGTGGGAGTGGCGC 245
Db      646 GlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysPheProGlyGlnGly 665
Qy      246 ATACACCGAGTTCCCTTGGCGCAGACAGCTGGGTGAGCGTCCAGAGATGCGGTGG 305

```

```

Db 666 Val-----ProGlyAspLeu-----GlyAlaProGlyProSerGly 677
Qy 306 TCCTTCGCCAGCCGATGAGTGAACCCACACTGCGCTGCTGATGCCAAGA 365
Db 678 -----AlaArgGly-----GluArgGlyPheProGlyGluArgGlyValGlnGly 692
Qy 366 ACAAGCTTTGCTGAAGCATCAAGTCCCGGCC-----AGGGGTGCCCAAAAGTCTGCT 419
Db 693 ProPro-----GlyProAlaGlyProArgGlyAlaAsnGlyAlaPro 706
Qy 420 GTGGGAGCAGATGAGCGGCTCCAGCGACCTGTGGGCCCCCCTATGCA---GGAGCTCAA 476
Db 707 GlyAsnAsp-----GlyAlaGlySerGlyAlaProGlyAlaProGlySerGln 724
Qy 477 GGA---CCTGGGCGTCAACGCTGCTATTCATGACACCGGCGGCACTTCTGGAGCT 533
Db 725 GlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyPro 744
Qy 534 GTACGCCGCTGCTCAGCCCTGCGGAGAAACACTGCACTTTGTGAGACGTGATCACT 593
Db 745 -----GlySerGlyAspArg 748
Qy 594 GCAACATCATTTCCAGA---GCTGAGGGGCTCCATTCTCGGATGCGGCC----- 641
Db 749 GlyAspAlaGlyProGlySerGlyAlaAspGlySerProGlyValArgGlyLeu 768
Qy 642 -----GGAGCAGCTCATGCAACCGAGATCACTGTCAG 674
Db 769 ThrGlyProGlyGlyProProGlyProAlaGlyAlaPro-----GlyAspGlySerGln 786
Qy 675 CGGCTTCGCGCTGCGCTGCGCACCTGCTGACCGGCACTCGGGCTACTATGCTGGA 734
Db 787 SerGlyProSerGlyProAlaGlyProThrGlyAlaArg-----Gly 800
Qy 735 GCTGTCGCCAGCGCCAGCGGCGCTGCAAGACCCGACGACCTGCGGAAACCCAC 794
Db 801 AlaProGlyAspArgGlyGluProGlyProGlyProAlaGlyPheAlaGlyProPro 820
Qy 795 GGAAGTGAATCTGGCGCGCT-----CGACCGGACACGAGACTACGACGTCGC 842
Db 821 GlyAlaAsp-----GlyGlnProGlyAlaGlySerGlyGluProGlyAlaAspAlaGlySerGly 839
Qy 843 GCTAGTGCCTGAGTCCACGCTGCGCTCTGAGCGCCAGATCTTGC 890
Db 840 -----AspAlaGlyProProGlyProAlaGlyProAlaGlyProProGly 854
Qy 891 -----GCTGCGCACGCGCGCCAGAGGAGCGCG----- 917
Db 855 ProGlyGlyAsnValGlyAlaProGlyAlaGlyAlaArgGlySerAlaGlyProPro 874
Qy 918 -----GCCAGAGCGCATGTCATCTCCACGCGCGGCC----- 950
Db 875 GlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsn 894
Qy 951 GCGCAGACCTTCGCGTGAAGTTGGGCGCCAGCGCTGAGCTGACCGCGCGCTGACCA 1010
Db 895 AlaGlyProProGlyProProGlyProAlaGly-----LysGlnGlyGlySerGlyPro 912
Qy 1011 CGTGCAAGTTCGGGCGCTGCGGCGGCGGAGCGCAGCGGT----- 1052
Db 913 ArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProProGlyProProGly 932
Qy 1053 -----GGAGGTGCGCGCGCGCGCACTGACACCGCTGACAGG 1091
Db 933 ProAlaGlyGlyLeuSerProGlyAlaAspGlyPro-----AlaGlyAla 948
Qy 1092 CCTGAGC-----GCGGAGCAGCGC-----CTACCTGCT 1118
Db 949 ProGlyThrProGlyProGlnGlyLeuAlaGlyAlaArgGlyValAlaGlyLeuProGly 968
Qy 1119 GACCGTGAACCGCGC-----CTTCGCGCTGCGCGCGGAGAGCGCGCTGTCGC 1166
Db 969 GlnArgGlyGluArgGlyPheProGlyLeuPro-----GlyProSerGlyGluProGlyLeu 987

```

```

Qy 1167 CAAGCCTGACAGCCCGACGCGCCGCGCCAGCCCGCTGCGCCCGCCAC 1226
Db 988 GlnGly-----ProSerGlyAlaSerGlyGluArgGlyPro--ProGlyP 1002
Qy 1227 CCGCGGAGCCCGACCGTGAGC 1249
Db 1002 romerGlyProProGlyLeuAla 1009

RESULT 14
US-11-000-463-243
; Sequence 243, Application US/11000463
; Publication No. US2005026423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radjoe T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIPACN
; CURRENT APPLICATION NUMBER: US/11/000,463
; PRIOR FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-000-463-243

Alignment Scores:
Score: 9.71e-06 Length: 1464
Percent Similarity: 211.00 Matches: 128
Best Local Similarity: 31.2% Conservative: 30
Query Match: 8.9% Mismatches: 180
DB: 7 Gaps: 30

US-10-699-035A-5 (1-1254) x US-11-000-463-243 (1-1464)
Qy 1243 GGCTGCGGTCCCGGAGTCCGGGCGCGGAGCGGAGCGGCGCGGCGCT 1184
Db 242 GlyGluArgGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 261
Qy 1183 CGGCGTGAAGCGCTTGGCGG-----ACAGCGCC 1154
Db 262 ProGlyMetLeuGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAspAlaGly 281
Qy 1153 TCTCGGCGCGCGGAGGCGGAGGCGGCGTCA-----CGGTCAACAGGTAGCGG 1106
Db 282 ProAlaGlyProLysGlyGluProGlySerProGlyGluAsnGlyAlaProGlyGlnMet 301

```

```

QY 1105 TGCCCG-----GCCGACAGCCCTGCA----- 1085
Db 302 GlyProArgGlyLeuProGlyGlyLuarGlyArgProGlyAlaProGlyProAlaGlyLys 321
QY 1084 -----GCCGTGGTGCAGTTGGCGCCGCGGACCTCCA-----CCCGTGGG 1043
Db 322 ArgGlyAsnAspGlyAlaThrGlyAlaAlaGlyProProGlyProThrGlyProAlaGly 341
QY 1042 CCTCCCGCCCGCCGACGCGGCAACTGCACTGTGAGCCGAGCGCGCGCTGAGCCCA 983
Db 342 ProProGlyPheProGlyAlaValaGlyAlaGlyGlyAlaGlyProIn----- 358
QY 982 GCGCTGGGCGCCCACTCAGCGGAGGCTCGCGCGCGCGGTGGAGATGACGATCGCT 923
Db 359 -----GlyProArgGlySerGlyGlyProGlyAlaGly----- 371
QY 922 CTGGCGCCGCGCT-----CCTCGCGCGGTGGCACTCCGACGATCT---GGG 878
Db 372 -----GluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnProGly 389
QY 877 GCGTCAGGA-----GCGCGACGTTGGACTCAGGCACTAGCGCCAGCTGTGCTGCT 824
Db 390 AlaAspGlyGlnProGlyAlaLysGlyAlaAsnGly---AlaProGlyLeuAlaGlyAla 408
QY 823 CCGGGT----- 818
Db 409 ProGlyPheProGlyAlaArgGlyProSerGlyProGlyGlyProGlyGlyProProGly 428
QY 817 -----CGAGCGCGCGCCAGATCCAGTCCGTCGTCCTCCGACGCTGCT--- 773
Db 429 ProLysGlyAsnSerGlyGlyProGlyAlaProGlySerGlyAspThrGlyAlaLys 448
QY 772 -----GCGCTTGTGACCCCGC----- 755
Db 449 GlyGluProGlyProValaGlyAlaGlyGlyProProGlyProAlaGlyGluGlyLys 468
QY 754 ---GCTGGCGCTGGGACACGACTCTAGCA-----CATACT 722
Db 469 ArgGlyAlaArgGlyGlyLuarProGlyProThrGlyLeuProGlyProProGlyGlyLuarGly 488
QY 721 AGCCGAGCTGCGGTC-----GCGGGGTGGCCAGGCCAGGC 683
Db 489 GlyProGlySerArgGlyPheProGlyAlaAspGlyAlaGlyProLysGlyProAla 508
QY 682 GGAAGCGCGTGGACGATCTCCGTGGACAGCTGCGCGCGCATCGGCAAGATAG 623
Db 509 GlyLuarG-----GlySerProGlyProAlaGlyProLysGly 521
QY 622 AGCCCTCAGCTTGGACAATGATGTCAGGTCATCCAGCTCCACAAGTGCAGGTGCT 563
Db 522 SerProGlyGlyAlaGly-----ArgProGlyGlyAlaGlyLeu 534
QY 562 TCTCGGACAGGCGTGCAGAGCGGCTGACAGCTCCAGAAAGTTGCTCGGCGGTCTGA 503
Db 535 -----ProGlyAlaLysGlyLeuThrGlySerProGlySer----- 546
QY 502 CAATGAACACGCTGACGCCAGGCTCTGAGCTCTGCAATG---GGGGCCCAAGGCT 446
Db 547 -----ProGlyProAspGlyLysThrGlyProProGlyProAlaGly 560
QY 445 CGCTGAGCGCGCATATGTCACCCACACAGCACTTTGGGACCCCTGGCCGGCACTGG 386
Db 561 GlnAspGlyArgProGlyProProGlyAlaArgGlyGlnAlaGlyValMet 580
QY 385 ATGCTTCACAAACACTGTTCTTGGCAATAGACAGCGCGCAAGTGGGTGTAC 326
Db 581 -----GlyPheProGlyProLysGlyAlaAlaGlyLuarProGly----- 593
QY 325 CCATGCGCTGGGAGAGACGACCGCATCT----- 293
Db 594 ---LysAlaGlyGlyLuarGlyValProGlyProProGlyAlaValaGlyProAlaGlyLys 612
QY 292 -----GGGACGCTCACCAGC---TGTGCTGCGCGAAGGGA 257

```

```

Db 613 AspGlyLuarAlaGlyAlaGlyGlyProProGlyProAlaGlyProAlaGlyLuarGly 632
QY 256 ACTCGGTATGCGCGCATGCTC-----CCAGTGCACCAAC 221
Db 633 GlnGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaGlyProPro 652
QY 220 TGGACGAGAGGCGCCCGGTGCGCGACGTGAGCAACCACTGCCCCCAAACTCC 161
Db 653 GlyLuarAlaGlyLys-----ProGlyGluGlnGlyValProGlyAspLeuGlyAlaPro 670
QY 160 GAACCCGGGAGAACTCGTAGAGAGA-----CGTGGCTGACGTGTCCA 116
Db 671 GlyProSerGlyAlaArgGlyGlyLuarGlyPheProGlyGlyLuarGlyValGlnGlyPro 690
QY 115 GCAGGAATCAGATGCTCCCTCGGGGCGTGAATGCTGTGACCGCGCTCCGCGCGCTCC 56
Db 691 ProGlyProAlaGlyProArgGly-----AlaAsnGlyAlaProGlyAsnAsp 706
QY 55 GCGCCAGCGCCAGCCGCAAGCTCAGGCGCAGCGCGCGCTCCAGGAGCA 2
Db 707 Gly---AlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGlnGlyAla 723

RESULT 15
US-11-186-284-28
/ Sequence 28, Application US/11186284
/ Publication No. US2005026493A1
/ GENERAL INFORMATION:
/ APPLICANT: Millennium Pharmaceuticals, Inc.
/ APPLICANT: Berger, Allison
/ APPLICANT: Guillemette, Tracy L.
/ APPLICANT: Kamatkar, Shubhangi
/ APPLICANT: Schlegel, Robert
/ APPLICANT: Monahan, John E.
/ APPLICANT: Thibodeau, Stephen N.
/ APPLICANT: Burgart, Lawrence J.
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
/ TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
/ FILE REFERENCE: MPM01-029P2RNM
/ CURRENT APPLICATION NUMBER: US/11/186,284
/ CURRENT FILING DATE: 2005-07-21
/ PRIOR APPLICATION NUMBER: US/10/301,822
/ PRIOR FILING DATE: 2002-11-21
/ PRIOR APPLICATION NUMBER: US 60/339,971
/ PRIOR FILING DATE: 2001-12-10
/ PRIOR APPLICATION NUMBER: US 60/361,978
/ PRIOR FILING DATE: 2002-03-05
/ PRIOR APPLICATION NUMBER: US 60/381,988
/ NUMBER OF SEQ ID NOS: 228
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 28
/ LENGTH: 1464
/ TYPE: PRT
/ ORGANISM: Homo Sapiens
US-11-186-284-28

Alignment Scores:
Pred. No.: 9 71e-06 Length: 1464
Score: 211.00 Matches: 138
Percent Similarity: 31.2% Conservative: 30
Best Local Similarity: 25.7% Mismatches: 190
Query Match: 8.9% Indels: 180
DB: 7 Gaps: 30

US-10-699-035A-5 (1-1254) x US-11-186-284-28 (1-1464)
QY 1243 GCGTGGCGGTCCCGCGGCGCGGAGCAGCGGCGGTGGCGCGCGGCGCGGCGCT 1184
Db 242 GlyLuarArgGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 261
QY 1183 CGGCGGTGACGCGCTTGGCGG-----ACAGCGCC 1154

```


THIS PAGE BLANK (USPTO)


```

DB:                2                Gaps:                2
US-10-699-035A-1 (1-537) x S31212 (1-1857)
OY      4 GACCTGATGTTCTGCTGAGACGCTCAGACGAGCTCTCTCACTAGAGTTCTCCGGGTT 63
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      158 AsplleValIleLeuValAspGlySerTrpSerIleGlyArgPheAsnPhenArgLeuVal 177
OY      64 CGGAGTTTGTGGGCGAGCTGTGGCTGCACGCTCCAGCCCTGGAGCGGGGCGCTGCGCC 123
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      178 ArgLeuPheLeuGluAsnLeuValSerAlaPheAsnValGlySerGluLysThrArgVal 197
OY      124 AGCTGTGTGCACGTGGGCGAGCTGCGCATACACGAGTTCCCTTGGCGGACGACGCTCG 183
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      198 GlyLeuAlaGlnIlyrSerGlyAspProArgIleGluTrpHisLeuAsnAlaIlyrGlyThr 217
OY      184 GGTGAGGCTGCCCGACGATGCGGTGCTTGTGCGCCAGCGGATGGGTGACACCACT 243
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      218 LysAspAlaValIleuAspAlaValAlaArgAsnLeuProIlyrGlyAsnThrLeuThr 237
OY      244 GGCCTGTGGCGCTGTCTATGCTCAAGAAACAGCTGTTTGTGAAGCATCAGTGCCTCCGCGCA 303
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      238 GlyLeuAlaLeuThrTyrIleLeuGluAsnSerPheLysProGluAlaGlyAlaArgPro 257
OY      304 GGGGTGCCCAAGTGTGTGTGGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCGCC 363
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      258 GlyValSerIlyrIleGlyIleLeuIleThrAspGlyLysSerGlnAspValIlePro 277
OY      364 CCCATGCGAGAGCTCAAGAGACCTGGCGCTCACCGGTTCATTGTGACGACGCGCGAGGCG 423
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      278 ProAlaLysAsnLeuArgAspAlaGlyIleGluLeuPheAlaIleGlyValLysAsnAla 297
OY      424 AACTTCCTGAGAGCTGTGCAGCCGCTGCTCAGCCCTGCGCGAGAGACCTGTCGACTTGTG 483
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      298 AsplleAsnGluLeuLysGluIleAlaSerGluProAspSerThrHisValIlyrAsnVal 317
OY      484 ---GACGTGATGACCTGCACATCATTTGCCA----- 513
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      318 AlaAspPheAsnPhenMetAsnSerIleValIleGluGlyLeuThrArgThrValCysSerArg 337
OY      514 -----GAGCTGAGGGGCTCCATTCTCGCG 537
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      338 ValGluGluGlnGluLysGluIleLysGlyIleAlaAla 351
Db
RESULT 2
S78476
collagen alpha 1(XIV) chain precursor, long form - chicken
C:Species: Gallus gallus (chicken)
C>Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 09-Jul-2004
C:Accession: S78476; S31211
R:Trueb, B.
submitted to the EMBL Data Library, January 1993
A:Reference number: S78476
A:Accession: S78476
A:Molecule type: mRNA
A:Residues: 1-1888 <TRU>
A:Cross-references: UNIPROT:P32018; UNIPARC:UPI0000126D31; EMBL:X70793; NID:G288872; PID
R:Waelechi, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A:Title: Complete primary structure of chicken collagen XIV.
A:Reference number: S31211; MUID:93185668; PMID:8444186
A:Molecule type: mRNA
A:Cross-references: UNIPARC:UPI00001773F4; UNIPARC:UPI00001773F5; EMBL:X70793
C:Genetics:
A:Gene: Col14A1
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F:1-28/Domain: signal sequence #status predicted <Sig>
F:29-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <Mat>
F:29-110/Domain: fibronectin type III repeat homology <FN3A>
F:156-320/Domain: von Willebrand factor type A repeat homology <VWA1>

```

```

F:352-433/Domain: fibronectin type III repeat homology <FN3B>
F:442-525/Domain: fibronectin type III repeat homology <FN3C>
F:534-614/Domain: fibronectin type III repeat homology <FN3D>
F:623-707/Domain: fibronectin type III repeat homology <FN3E>
F:741-823/Domain: fibronectin type III repeat homology <FN3F>
F:832-914/Domain: fibronectin type III repeat homology <FN3G>
F:922-1009/Domain: fibronectin type III repeat homology <FN3H>
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Alignment Scores:
Pred. No.:      1,15e-13      Length:      1888
Score:          268.00      Matches:      64
Percent Similarity: 50.5%      Conservative: 34
Best Local Similarity: 25.0%      Mismatches: 80
Query Match:    33.5%      Indels:      16
DB:             2                Gaps:      2

US-10-699-035A-1 (1-537) x S78476 (1-1888)
OY      4 GACCTGATGTTCTGCTGAGACGCTCAGACGCTCTCTCACTAGAGTTCTCCGGGTT 63
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      158 AsplleValIleLeuValAspGlySerTrpSerIleGlyArgPheAsnPhenArgLeuVal 177
OY      64 CGGAGTTTGTGGGCGAGCTGTGGCTGCACGCTCCAGCCCTGGAGCGGGGCGCTGCGCC 123
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      178 ArgLeuPheLeuGluAsnLeuValSerAlaPheAsnValGlySerGluLysThrArgVal 197
OY      124 AGCTGTGTGCACGTGGGCGAGCTGCGCATACACGAGTTCCCTTGGCGGACGACGCTCG 183
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      198 GlyLeuAlaGlnIlyrSerGlyAspProArgIleGluTrpHisLeuAsnAlaIlyrGlyThr 217
OY      184 GGTGAGGCTGCCCGACGATGCGGTGCTTGTGCGCCAGCGGATGGGTGACACCACT 243
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      218 LysAspAlaValIleuAspAlaValAlaArgAsnLeuProIlyrGlyGlyAsnThrLeuThr 237
OY      244 GGCCTGTGGCGCTGTCTATGCTCAAGAAACAGCTGTTTGTGAAGCATCAGTGCCTCCGCGCA 303
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      238 GlyLeuAlaLeuThrTyrIleLeuGluAsnSerPheLysProGluAlaGlyAlaArgPro 257
OY      304 GGGGTGCCCAAGTGTGTGTGGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCGCC 363
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      258 GlyValSerIlyrIleGlyIleLeuIleThrAspGlyLysSerGlnAspValIlePro 277
OY      364 CCCATGCGAGAGCTCAAGAGACCTGGCGCTCACCGGTTCATTGTGACGACGCGCGAGGCG 423
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      278 ProAlaLysAsnLeuArgAspAlaGlyIleGluLeuPheAlaIleGlyValLysAsnAla 297
OY      424 AACTTCCTGAGAGCTGTGCAGCCGCTGCTCAGCCCTGCGCGAGAGACCTGTCGACTTGTG 483
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      298 AsplleAsnGluLeuLysGluIleAlaSerGluProAspSerThrHisValIlyrAsnVal 317
OY      484 ---GACGTGATGACCTGCACATCATTTGCCA----- 513
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      318 AlaAspPheAsnPhenMetAsnSerIleValIleGluGlyLeuThrArgThrValCysSerArg 337
OY      514 -----GAGCTGAGGGGCTCCATTCTCGCG 537
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      338 ValGluGluGlnGluLysGluIleLysGlyIleAlaAla 351
Db
RESULT 3
A40020
collagen alpha 1(XII) chain precursor - chicken
N:Alternate names: fibrochimerin
C:Species: Gallus gallus (chicken)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811
R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obari
J. Cell Biol. 115, 209-221, 1991
A:Title: The complete primary structure of type XII collagen shows a chimeric molecule w
nous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site
A:Reference number: A40020; MUID:92011862; PMID:1918137
A:Molecule type: mRNA

```

A:Residues: 1-3124 <YAM>
A:Cross-references: UNIPROT:P13944; UNIPARC:UPI0000126D2D; GB:D00824; NID:g222810; PIDN:
A>Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and
R.Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
J. Biol. Chem. 264, 19772-19778, 1989
A>Title: Type XII collagen. A large multidomain molecule with partial homology to type I
A:Reference number: A34485; MUID:90062079; PMID:2584192
A:Molecule type: A34485
A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'P', 2978-3124 <GOR>
A:Cross-references: UNIPARC:UPI0000171233; EMBL:J05137; NID:g211284; PIDN:AAA48635.1; PM
A:Accession: B34485
A:Molecule type: protein
A:Residues: 2772-2792; 2846-2873 <GOR2>
A:Cross-references: UNIPARC:UPI0000173C3F; UNIPARC:UPI0000173C40
R.Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A>Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA
A:Reference number: A28037; MUID:87317590; PMID:3476925
A:Molecule type: mRNA
A:Residues: 2960-2976, 'F', 2978-3074, 'AG' <GOR3>
A:Cross-references: UNIPARC:UPI0000172F4; EMBL:M17375; NID:g211649; PIDN:AAA48718.1; PM
A>Note: This sequence has been revised in reference A34485
R.Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992
A>Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A:Reference number: S23814; MUID:92362621; PMID:1323460
A:Accession: S23814
A:Molecule type: protein
A:Residues: 'X', 1333, 'Q', 1335-1347; 1914-1928; 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517
A:Cross-references: UNIPARC:UPI0000173C41; UNIPARC:UPI0000173C42; UNIPARC:UPI0000173C43
R.Dublet, B.; van der Rest, M.
J. Biol. Chem. 262, 17724-17727, 1987
A>Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-
A:Reference number: S22254; MUID:88087065; PMID:3121603
A:Accession: S22254
A:Molecule type: protein
A:Residues: 2831-2832, 'T', 2834, 'R', 2836-2843; 3002-3014 <DUH>
A:Cross-references: UNIPARC:UPI0000173C44; UNIPARC:UPI0000173C45
R.Trueb, J.; Trueb, B.
Biochim. Biophys. Acta 1171, 97-98, 1992
A>Title: The two splice variants of collagen XII share a common 5' end.
A:Reference number: S28811; MUID:93042014; PMID:1420368
A:Accession: S28811
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>
A:Cross-references: UNIPARC:UPI0000173C46; EMBL:X67327
C:Genetics: S4
A:Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
C:Keywords: alternative splicing; cell binding; coll; connective tissue; disulfid
F:1/3-Domain: signal sequence #status predicted <SIG>
F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
F:24, 1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted
F:24-114/Domain: IIIA #status predicted <IIIA>
F:24-105/Domain: fibronectin type III repeat homology <FN3A>
F:137-301/Domain: von Willebrand factor type A repeat homology <VWA1>
F:332-425/Domain: IIIB #status predicted <IIIB>
F:332-414/Domain: fibronectin type III repeat homology <FN3B>
F:437-601/Domain: von Willebrand factor type A repeat homology <VWA2>
F:629-1178/Domain: IIIC #status predicted <IIIC>
F:630-711/Domain: fibronectin type III repeat homology <FN3C>
F:721-802/Domain: fibronectin type III repeat homology <FN3D>
F:812-895/Domain: fibronectin type III repeat homology <FN3E>
F:905-986/Domain: fibronectin type III repeat homology <FN3F>
F:995-1076/Domain: fibronectin type III repeat homology <FN3G>
F:1086-1169/Domain: fibronectin type III repeat homology <FN3H>
F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>
F:1384-2295/Domain: IID #status predicted <IID>
F:1474-1557/Domain: fibronectin type III repeat homology <FN3I>
F:1566-1647/Domain: fibronectin type III repeat homology <FN3K>

[illegible]

Db 82 G1YleAlaIagIntYserGIYasPProArgI1eGIuTrpH1sleuAsnAlaTyrcIcIYthr 101

QY 184 GGTGAGCTGCCAGATGCGGTGCTGCTTTCGCCAGCGCATGGGTGAACCCACT 243

Db 102 LysAspAlaValleuAspAlaAspArgAsnLeuProTyrIYsgIcIYAsnThrIeuThr 121

QY 244 GGCCCTGGCGCGCTGTCATATGCCAAGAACAGCTGTTTCTGTGAAGCATCAGTGGCCGGCA 303

Db 122 G1YleAlaIeAuThrTyrI1leuGIuAsnCysPheYsPProG1uAlaGIYAlaArgPro 141

QY 304 GGGGTGCCCAAGTCTGGTGGTGGTGACAGATGCGCGCTCCAGCACTCTGGGCCCC 363

Db 142 G1YValserIYleIleGIYleuIleThrAspGIYsSerGIYAsnAspValIlePro 161

QY 364 CCCATCAGAGAGCTCAGACGACTGGCGCTGACCCCTGTTTCATTGTCAGACCGCGGAGGC 423

Db 162 ProAlaIYAsnIeuArgIuAlaGIYIleGIuIeuPheAlaIeGIYValIYAsnAla 181

QY 424 AACTTCCTGGAGGTGCACAGCGCGCTCAGCCCTCCGAGAACCTCAGCTGTTGG 483

Db 182 AspI1eAsnGIuIeuYsgIuIeAlaSerGIuProAspSerThrIshValIYAsnVal 201

QY 484 ---GAGCTGATGATCCTGCATCATTTGTCGA----- 513

Db 202 AlaAspPheAsnPheMetAsnSerIleValGIuGIuIeuThrArgThrValCysSerArg 221

QY 514 -----GAGCTGAGCGGCTCCATTCTCGCG 537

Db 222 ValGIuGIuGIuGIuIYsgIuIeYsgIYThrIleAlaIa 235

RESULT 5

AJ3809

cartilage matrix protein precursor - chicken

C1Species: Gallus gallus (chicken)

C1Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 09-Jul-2004

C1Accession: AJ3809; A26364

R1Kiss, I.; Deak, F.; Holloway Jr., R.G.; Delius, H.; Meubst, K.A.; Frimberger, E.; Argir

U. Biol. Chem. 264, 8126-8134, 1989

A1Title: Structure of the gene for cartilage matrix protein, a modular protein of the ex

grins, Von Willebrand factor, complement factors B and C2, and epidermal growth factor.

A1Reference number: AJ3809; MUID:89255246; PMID:2542265

A1Accession: AJ3809

A1Status: preliminary

A1Molecule type: DNA

A1Residues: 1-493 <RIS>

A1Cross-references: UNIPROT:P05099; UNIPARC:UPI000004PIEF; GB:X12346; GB:X12347; GB:X123

Proc. Natl. Acad. Sci. U.S.A. 84, 464-468, 1987

R1Argaves, W.S.; Deak, F.; Sparks, K.J.; Kiss, I.; Goetlinck, P.F.

A1Title: Structural features of cartilage matrix protein deduced from cDNA.

A1Reference number: A26364; MUID:87092429; PMID:3025875

A1Accession: A26364

A1Molecule type: mRNA

A1Residues: 78-493 <ARG>

A1Cross-references: UNIPARC:UPI000004PIEF; GB:M14792; NID:9211545; PIDN:AAA48695.1; PID:

C1Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe

F137-204/Domain: von Willebrand factor type A repeat homology <VWA1>

F1225-260/Domain: EGF homology <EGF>

F1270-434/Domain: von Willebrand factor type A repeat homology <VWA2>

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
5.36e-12	247.50	493	62
Percent Similarity:	51.5%	Conservative:	22
Best Local Similarity:	38.0%	Mismatches:	78
Query Match:	23.6%	Indels:	1
Db:	2	Gaps:	1

US-10-699-035A-1 (1-537) x AJ3809 (1-493)

QY 4 GACCTGATGTTCTGCTGACAGCTCAGCCAGCGTCTCACTACGAGTTCTCCGGTT 63

Db 272 AspLeuValPheIeuIleAspGIYserIYsSerValArgProGIuAsnPheGIuIeuVal 291

[illegible]

RESULT 8
A:Accession: A54849
N:Collagen alpha 1(VII) chain precursor - human
N:Alternate names: procollagen alpha 1(VII) chain
C:Species: Homo sapiens (man)
C:Date: 04-Nov-1994 #sequence revision 04-Nov-1994 #text change 09-Jul-2004
C:Accession: A54849; PH0844; S16316; I56328; A10296; I64686
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Altobelli, J.
J. Biol. Chem. 269, 20256-20262, 1994
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha(VII)
A:Reference number: A54849; MUID:94327588; PMID:8051117
A:Accession: A54849
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2944 <CHR>
A:Cross-references: UNIPROT:002388; UNIPARC:UPI000017A138; GB:L02870; NID:G987124; PIDN:
R:Tanaka, T.; Takahashi, K.; Fukukawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A:Title: Molecular cloning and characterization of type VII collagen cDNA.
A:Reference number: PH0844; MUID:92231902; PMID:1567409
A:Accession: PH0844
A:Molecule type: mRNA
A:Residues: 'ERR', 340-475, 'RALSTRASHSTLCWRATRWHPGCRSGHMTFRACEPCNRPASHRAANAG', 524-528, 'C', 'C'
A:Cross-references: UNIPARC:UPI000017A139; DDBJ:D11152; DDBJ:D13694; NID:G453698; PIDN:
A:Experimental source: Keratinocyte
A:Note: The authors translated the codon ACC for residues 394 and 397 as Tyr
R:Pariente, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A:Reference number: S16316; MUID:91334580; PMID:1871109
A:Accession: S16316
A:Molecule type: mRNA
A:Residues: 815-892, 'E', 894-1439 <PAR>
A:Cross-references: UNIPARC:UPI000016A722; GB:M65158; GB:S49017; NID:g180914; PIDN:AA9
A:Experimental source: Keratinocyte
R:Gammon, W.R.; Aderchty, M.L.; Padilla, K.M.; Prisayanb, P.S.; Cook, M.E.; Wright, J.
J. Invest. Dermatol. 99, 691-696, 1992

A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion protein
A:Reference number: 156328; MUID:93107742; PMID:1469284
A:Accession: 156328
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Cross-references: EPR,372-517,1DV,520-540,W,542-1255 <RES>
A:Cross-references: UNIPARC:UPI000016B3AC; GB:551236; NID:g362308; PION:AB42637.1; PID:R.Saltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glenville, R.W.; Burgeson, R.E.
U. Biol. Chem. 264, 3822-3826, 1989
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase
A:Reference number: A30296; MUID:89139437; PMID:2537292
A:Accession: A30296
A:Molecule type: protein
A:Residues: 'A',1240-1246,'G',1248-1250,'XE',1253-1255,'O',1257,'E',2032,'C',2034-2041,'
A:Cross-references: UNIPARC:UPI000017A13A; UNIPARC:UPI0000017A13B; UNIPARC:UPI000017A13C;
A:Note: two reported peptides cannot be reliably located
R:Greenepan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A:Reference number: 148103; MUID:93271985; PMID:8499916
A:Accession: 184686
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 2395-2871,'S',2873-2944 <RE2>
A:Cross-references: UNIPARC:UPI000016A724; GB:106662; NID:g388713; PION:AAA89196.1; PID:
R.Christiano, A.M.; Rynaenen, M.; Utico, U.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser substitution
A:Reference number: A55255; MUID:94224777; PMID:8170945
A:Contents: annotation
A:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (repeated and subsequently O-glycosylated).
C:Genetics:
A:Gene: GDB:COL7A1; EBR1; EBD1; EB
A:Cross-references: GDB:128750; OMIM:120120
A:Map position: 3p21.3-3p21.3
A:Note: detects in this gene can result in dominant and recessive dystrophic epidermolysis
A:Note: there are 118 introns
C:Complex: type VII collagen is probably a homotrimer
C:Function:
A:Description: structural component of extracellular polymer associated with anchoring fibrils
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F:36-201/Domain: von Willebrand factor type A repeat homology <VWA1>
F:231-318/Domain: fibronectin type III repeat homology <FN1>
F:327-413/Domain: fibronectin type III repeat homology <FN2>
F:414-502/Domain: fibronectin type III repeat homology <FN3>
F:508-593/Domain: fibronectin type III repeat homology <FN4>
F:598-683/Domain: fibronectin type III repeat homology <FN5>
F:686-771/Domain: fibronectin type III repeat homology <FN6>
F:776-862/Domain: fibronectin type III repeat homology <FN7>
F:864-952/Domain: fibronectin type III repeat homology <FN8>
F:954-1045/Domain: fibronectin type III repeat homology <FN9>
F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>
F:1170-1172/Region: cell attachment (R-G-D) motif
F:1189-1253/Region: cysteine/proline-rich
F:1254-3783/Region: interrupted helical
F:1334-1336/Region: cell attachment (R-G-D) motif
F:2008-2010/Region: cell attachment (R-G-D) motif
F:2553-2555/Region: cell attachment (R-G-D) motif
F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F:3287-3929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F:337,786,1109/Binding site: carboxylate (Asn) (covalent) #status predicted
F:2187,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status experimental
F:2653,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F:2653,2631/Binding site: carboxylate (Lys) (covalent) #status experimental
F:2634,2802,2804/Disulfide bonds: interchain #status predicted

QY 475 CACTTTGGACGATGATGACCTGCAATCTGTCACGAGGCTGAGGGCTCCATT 531
Db 198 AsnLeuGluAenPheThrAlaLeuHISglYleValglAspLeuValAlaSerVal 216
RESULT 10
T46488
hypothetical protein DKFZp434J065.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C/Accession: T46488
R/Diesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A/Reference number: Z23035
A/Accession: T46488
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-741 <AAA>
A/Cross-references: UNIPROT:Q8NDE6; UNIPARC:UPI000016ACCB; EMBL:AL137638
A/Experimental source: adult testis; clone DKFZp434J065
C/Genetics:
A/Note: DKFZp434J065.1
Alignment Scores:
Pred. No.: 2.97e-09 Length: 741
Score: 213.00 Matches: 58
Percent Similarity: 50.8% Conservative: 34
Best Local Similarity: 32.0% Mismatches: 83
Query Match: 20.3% Indels: 6
Gaps: 4
US-10-699-035a-1 (1-537) x T46488 (1-741)
QY 4 GACCTGATGTTCTGCTGGACAGCTACGCGTCTCTCACTACGAGTCTCCGGGTT 63
Db 440 AspLeuValPheValIleAspGlySerIleuGluGluAenPheGluValVal 459
QY 64 CGGAGATTGTGGGGACAGCTGGCTCCACTGCCCTGGGACCGGGCCCTGGCTCC 123
Db 460 LysGlnPheValIleThrGlyIleLeuAspSerLeuThrIleSerProIleAlaAlaGlyVal 479
QY 124 AGTGTGGACAGCTGGGACAGCTGGGACAGCTGGGACAGCTGGGACAGCTGG 183
Db 480 GlyLeuLeuGlnIYserThrGlnAlaHISThrGluPheThrLeuAlaAsnPheAsnSer 499
QY 184 GGTAGGCTGCCCGAGATGCGGATGCGCTCTTCTGCGGACCGCATGGGT---GACACCCAC 240
Db 500 AlaIysAspMetIysIysAlaVal---AlaHISmetIysIYrMetGlyIYsGlySerMet 518
QY 241 ACTGAGCTGGCGCTGTCTATGCCAAGACAGCTGTTGTGAAGCATCAAGTCCCGG 300
Db 519 ThrGluLeuAlaLeuIleuHISmerPheGluIYrSerPheThrGlnGlyIYsAlaIYrG 538
QY 301 CCA-----GGGTCGCCCAAGTGTGTGTGGTGGGACAGATGGGCGCTCCAGCAC 351
Db 539 ProLeuSerThrAlaGlyAlaProAlaIleAlaIleValPheThrAspGlyAlaAlaIYrAsp 558
QY 352 CCTGTGGGCCCCCATGACAGACAGCTGGGACAGCTGGGACAGCTGGTGTGATTGTGACG 411
Db 559 AspValSerGluTyrPalaSerIysAlaIysAlaIYrIleThrMetYrAlaValIYs 578
QY 412 ACCGGCCGAGGCAACTTCTGAGCTGTACGCGCTGCTCAAGCCCTGCGGAGAGAC 471
Db 579 ValGluIYsAlaIleGluGluIYsAlaIleGluIYsAlaIleGluIYsAlaIleGluIYs 598
QY 472 CTGCACTTTGTG---GAGCTGATACCTGCACATCAATGTCGCAAGAGCTGAGGGGCTCC 528
Db 599 LeuPheTyrAlaGluIYsPheSerThrMetAspGluIYsSerGluIYsLeuIYsIYsGly 618
QY 529 ATT 531
Db 619 Ile 619
RESULT 11

CGH03A
collagen alpha 3(VI) chain precursor [validated] - human
M/Contents: collagen alpha 3(VI) chain, splice form A9/N10(-)
C/Species: Homo sapiens (man)
C/Date: 21-Nov-1993 #sequence_revision 12-Nov-1999 #text_change 16-Aug-2004
C/Accession: A59140; S13679; S24465; A57083; S28776; S00245; C31952; C29848; S26510; S48
R/Chu, M.L.
submitted to GenBank, May 1998
A/Reference number: A59140
A/Accession: A59140
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-3176 <CHU>
A/Cross-references: UNIPROT:P12111; UNIPARC:UPI0000126DAF; GB:X52022; NID:93127925; PIDN:
R/Chu, M.L.; Zhang, R.Z.; Pan, T.; Stokes, D.; Conway, D.; Kuo, H.T.; Glanville, R.; May
EMBO J. 9, 385-393, 1990
A/Title: Mosaic structure of globular domains in the human type VI collagen alpha-3 chain
A/Reference number: S13679; MUID:90151612; PMID:1689238
A/Accession: S13679
A/Molecule type: mRNA
A/Residues: 1-30,237-313, 'CWM', 318-322, 'AR', 326-1815, 'FD', 1818-1819, 'ID', 1822-3176 <CH5>
A/Cross-references: UNIPARC:UPI0000173C14; EMBL:X52022; NID:93127925
A/Accession: S24465
A/Molecule type: protein
A/Residues: 574-585; 965-973, 'X', 975-976; 1306-1355; 1361-1377; 1381-1401; 1473-1506, 'X', 1508
-1962, 'X', 1964-1965; 2018-2037; 2374-2410; 2445-2459; 2466-2469, 'X', 2471-2474; 2504-2508, 'X',
A/Cross-references: UNIPARC:UPI0000173C15; UNIPARC:UPI0000173C16; UNIPARC:UPI0000173C17;
CIC: UNIPARC:UPI0000173C1D; UNIPARC:UPI0000173C1E; UNIPARC:UPI0000173C1F; UNIPARC:UPI000
10000173C25; UNIPARC:UPI0000173C26; UNIPARC:UPI0000173C27
R/Zanussi, S.; Dollana, R.; Segat, D.; Bonaldo, P.; Colomacchi, A.
J. Biol. Chem. 267, 24082-24089, 1992
A/Title: The human type VI collagen gene. mRNA and protein variants of the alpha3 chain
A/Reference number: S28776; MUID:93054780; PMID:1339440
A/Accession: A57083
A/Molecule type: DNA
A/Residues: 310-328 <ZAN>
A/Cross-references: UNIPARC:UPI0000173C28
A/Accession: S28776
A/Molecule type: mRNA
A/Residues: 32-126, 'AK', 129-136, 'LV', 138-236 <ZAZ>
A/Cross-references: UNIPARC:UPI0000173C28; GB:S49432; NID:9260296; PIDN:AA24261.1; PID:
R/Chu, M.L.; Mann, K.; Deutermann, R.; Pribula-Conway, D.; Hsu-Chen, C.C.; Bernard, M.P.;
Eur. J. Biochem. 168, 309-317, 1987
A/Title: Characterization of three constituent chains of collagen type VI by peptide seq
A/Reference number: S00126; MUID:88029444; PMID:3665927
A/Accession: S00245
A/Molecule type: mRNA, protein
A/Residues: 2024-2046; 2092-2156, 'R', 2203-2208, 'X', 2210-2211, 'X', 2213-2227, 2228-2251, 2314
A/Cross-references: UNIPARC:UPI000016A705; UNIPARC:UPI0000173C29; UNIPARC:UPI0000173C2A;
ID:91335034
A/Note: The mRNA portion of the sequence corresponds to residues 2092-2157
R/Chu, M.L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutermann, R.; Timpl, R.
J. Biol. Chem. 263, 18601-18606, 1988
A/Title: Amino acid sequence of the triple-helical domain of human collagen type VI.
A/Reference number: A31952; MUID:89066644; PMID:3198591
A/Accession: C31952
A/Molecule type: mRNA
A/Residues: 2038-2373 <CH4>
A/Cross-references: UNIPARC:UPI0000173C2E; GB:J04211; GB:M20778
A/Note: parts of this sequence were determined by protein sequencing
R/Well, D.; Mattei, M.G.; Passage, E.; Van Cong, N.; Pribula-Conway, D.; Mann, K.; Deutz
Am. J. Hum. Genet. 42, 435-445, 1988
A/Title: Cloning and chromosomal localization of human genes encoding the three chains o
A/Reference number: A29848; MUID:88161046; PMID:3348212
A/Accession: C29848
A/Molecule type: mRNA
A/Residues: 2092-2151 <WEI>
A/Cross-references: UNIPARC:UPI000016A720; GB:M27449; NID:9291919; PIDN:AA52057.1; PID:
A/Note: part of this sequence was determined by protein sequencing
R/Jander, R.; Rautenberg, J.; Glanville, R.W.
Eur. J. Biochem. 133, 39-46, 1983
A/Title: Further characterization of the three polypeptide chains of bovine and human sh

```

A:Reference number: 526506; MUID:83209648; PMID:6852033
A:Accession: S26510
A:Molecule type: protein
A:Residues: 'SALGVGVG', <JAN>
A:Cross-references: UNIPARC:UP10000173C2F
A:Note: this sequence cannot be reliably placed and probably represents the results from
R:Mayer, U.; Poeschl, E.; Nischt, R.; Specks, U.; Pan, T.C.; Chu, M.L.; Timpl, R.
Eur. J. Biochem. 225, 573-580, 1994
A:Title: Recombinant expression and properties of the Kunitz-type protease-inhibitor mod
A:Reference number: 548709; MUID:95045506; PMID:7525281
A:Accession: 548709
A:Molecule type: mRNA
A:Residues: 'MRAMIFFLACGRALAA', 3102-3176 <MAY>
A:Cross-references: UNIPARC:UP10000173C30
A:Note: engineered sequence to allow isolation of the Kunitz-type proteinase inhibitor h
R:Arnoux, B.; Merigean, K.; Saidjian, P.; Norris, F.; Bjorn, S.; Olsen, O.; Timpl, R.
submitted to the Brookhaven Protein Data Bank, August 1994
A:Reference number: 452812; PDB:1KXT
A:Contents: annotation: X-ray crystallography, 1.6 angstroms, residues 3106-3160
A:Note: engineered sequence expressed in Saccharomyces cerevisiae strain mc-663
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
sines are 5-hydroxylated and subsequently O-glycosylated.
C:Comment: The fibronectin type III repeat homology domain may be released during proces
C:Genetics:
A:Gene: GDB:COL6A3
A:Cross-references: GDB:119066; OMTM:120250
A:Map position: 2q37.3-2q37.3
C:Complex: type VI collagen is a heterotrimer of one alpha 1(VI) chain (see PIR:GCHUJA),
sociations among trimer amino- and carboxyl-terminal domains (with disulfide bonds) .
C:Function:
A:Description: structural component of extracellular tissue microfibrils associated with
C:Superfamily: collagen VI
C:Keywords: alternative splicing; blocked amino end; cell binding; coiled coil; extracel
F:1-25/Domains: signal sequence #status predicted <SIG>
F:26-3176/Product: collagen alpha 3 (VI) chain #status predicted <MAT>
F:26-2037/Domains: amino-terminal nonhelical #status predicted <ANH>
F:26-30, 237-3176/Product: collagen alpha 3 (VI) chain, splice form A9/N10 (-) #status pred
F:37-203/Domains: von Willebrand factor type A repeat homology <VW01>
F:240-405/Domains: von Willebrand factor type A repeat homology <VW02>
F:443-608/Domains: von Willebrand factor type A repeat homology <VW03>
F:637-802/Domains: von Willebrand factor type A repeat homology <VW04>
F:835-999/Domains: von Willebrand factor type A repeat homology <VW05>
F:1027-1191/Domains: von Willebrand factor type A repeat homology <VW06>
F:1231-1394/Domains: von Willebrand factor type A repeat homology <VW07>
F:1434-1599/Domains: von Willebrand factor type A repeat homology <VW08>
F:1637-1802/Domains: von Willebrand factor type A repeat homology <VW09>
F:1836-2005/Domains: von Willebrand factor type A repeat homology <VW10>
F:2038-2373/Region: interrupted helical
F:2040-2042/Region: cell attachment (R-G-D) motif
F:2167-2138/Region: cell attachment (R-G-D) motif
F:2148-2150/Region: cell attachment (R-G-D) motif
F:2154-2156/Region: cell attachment (R-G-D) motif
F:2370-2372/Region: cell attachment (R-G-D) motif
F:2374-3176/Domains: carboxyl-terminal nonhelical #status predicted <CNH>
F:2400-2571/Domains: von Willebrand factor type A repeat homology <VW11>
F:2617-2800/Domains: von Willebrand factor type A repeat homology <VW12>
F:2865-2986/Region: alanine/lysine/proline/threonine/valine-rich repeats
F:2987-3072/Domains: fibronectin type III repeat homology <FN3>
F:3111-3161/Domains: animal Kunitz-type proteinase inhibitor homology <BPI>
F:26/Modified site: pyroglutamate carboxylic acid (Gln) (in mature form) #status predicted
F:108, 116, 202, 251, 2079, 2231, 2558, 2677, 2861, 3036/Binding site: carbonyl (Asn) (covalent)
F:2087/Disulfide bonds: interchain #status predicted
F:2100, 2206, 2239, 2316, 2319/Modified site: 4-hydroxyproline (Pro) #status experimental
F:2103, 2209, 2212, 2322, 2337/Binding site: carboxylate (Lys) (covalent) #status experimental
F:2103, 2209, 2212, 2322, 2337/Modified site: 5-hydroxylysine (Lys) #status experimental
F:3111-3161, 3120-3144, 3136-3157/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.: 7.55e-09 Length: 3176
Score: 207.50 Matches: 55
Percent Similarity: 47.4% Conservative: 28
Near Local Similarity: 31.4% Mismatches: 89
Query Match: 19.8% Indels: 3

```

```

DB: 2 Gaps: 2
US-10-699-035A-1 (1-537) x CGHJ3A (1-3176)

OY 4 GACSTGATGTTCTGCTGGACAGCTCAGCCAGCTTCTCACTACAGATTCTCCGGGTT 63
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 39 Asp1Le1IePheValAAspSerSerTrpThr1IeGluGluGlnHisPheG1nLeuVal 58
OY 64 CGGAGATTGTGGGGGACAGTGTGCTCCATCTGCCCCGGGACCGGGGCGCTGGTGC 123
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 59 ArgGluPheLeuTyAAspValValIysSerLeuAlaValG1GluAAspPheHisPhe 78
OY 124 AGTGGGAGCAGTGGGACAGTCCGACATACACAGCTCCCTTCCGACAGACAGCTG 183
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 79 AlaLeuValGlnPheAsnGlyAsnProHisThrGluPheLeuLeuAsnThrTyArgThr 98
OY 184 GGTGAGGCTGCCAGATGCGAGTCCGATGCTTCTGCCGACGATGGTGCACACCACT 243
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 99 LysGlnGluValIeAsnSerHisIeSerAsnMetSerTy1IeG1GlyThrAsnGlnThr 118
OY 244 GCGCTGGCGCTGTCTATCCCAAGAAACAGCTGTTTGTGAAGCATCAGTGCCTGG 300
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 119 G1YysG1YleuG1Ytyr1IeMetGlnSerHisLeuThrTyAlaG1YserTrgAla 138
OY 301 ---CCAGGGGGGCCCAAGATGCTGTGTGGGTGACAGATGGCGGCTGCAGACGCTG 357
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 139 G1YAspG1YAlaProGlnValIleValAlaLeuThrAspG1YHisSerTyAspG1YLeu 158
OY 358 GGGCCCCCATGACAGAGACTCAAGACACTGGGGCGCTACCGTGTATTGTACAGACCG 417
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 159 AlaLeuProSerAlaGluLeuLeuLysSerAlaAspValAlaAsnValPheAla1IeG1YAlu 178
OY 418 CGAGCAACTTCTGAGCTGTACAGCTGTACCGGCTCCTCAGGCGCTGCAGAAAGACCTG 474
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 179 AspAlaAspGluG1YAlaLeuLysGluIleLeuAsSerGluProLeuAsnMetHisMetPhe 198
OY 475 CACTTGTGACGCTGATGACTGCACATCATCTTGTCACAGAGCTG 519
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 199 AsnLeuGluAsnPheThrSerLeuHisAspR1IeValG1YAsnLeu 213

RESULT 12
A55348
Integrin alpha-1 - chicken (fragment)
C/Species: Gallus gallus (chicken)
C/Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004
C/Accession: A55348
R/Kern, A.; Brilesewitz, R.; Bank, I.; Marcantonio, E.B.
J. Biol. Chem. 269, 22811-22816, 1994
A/Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1
A/Accession: A55348; MUID:94357930; PMID:7521332
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-272 <KER>
A/Cross-References: UNIPROT:O42094; UNIPARC:UPI000017A14E; GB:U10114
F/55-230/Domain: von Willebrand factor type A repeat homology <vwa2>

Alignment Scores:
Pred. No.: 2,84e-08 Length: 272
Score: 201.00 Matches: 67
Percent Similarity: 47.4% Conservative: 26
Best Local Similarity: 34.2% Mismatches: 79
Query Match: 19.2% Indels: 24
DB: 2 Gaps: 8

US-10-699-035A-1 (1-537) x A55348 (1-272)

OY 4 GACSTGATGTTCTGCTGGACAGCTCAGCCAGCTTCTCACTACAGATTCTCCGGGTT 63
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 57 Asp1Le1IePheValAAspG1YSerAsnSerIle-----TyProTrpG1uSerVal 74
OY 64 CGGAGATTGTGGGGGACAGCTGTGCTCCATCTGCCCCCGGAGACCGGGGCGCTGGTGC 123
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```
Dh 75 ThrAlaPheLeuAenSerLeuLeuArgAsnMetAspIleGlyProGlnGlnThrGlnVal 94
Qy 124 AGCTGTCGACAGTCGTGGCAGTCGCGCAGTTCACCCAGTTCGGCCAGCAGCAGCTCG 183
   ::::: ||||| ||||| ||||| ::::: |||||
Db 95 GlyIleValGlnTyrglyGlnThrValAlaHisGluPheTyrlLeuAsnThrTySerThr 114
Qy 184 GGTGAGCTGCCAGAGTCGGTGGCTGCTTGGCCAGCGCATGGT--GACACCCAC 240
   ||||| ||||| ||||| ||||| ||||| |||||
Db 115 ThrGluGlnValMetAspAlaAlaLeuArgIleArgGlnArgGlyGlyThrGlnThrMet 134
Qy 241 ACTGGCCGTGGGCGTGTCTATGCGCAAGAGACGCTGTTTGGAGCATGAGTCCCGG 300
   ::::: ||||| ||||| ||||| ||||| ||||| |||||
Db 135 ThrAlaLeuGlyIleAspThrAlaArgGluGlnAlaPheThrGlnAlaHisGlyAlaArg 154
Qy 301 CCAGGAGTCCCAAGTGTGTGTGAGTGCAGATGGCGGCTCCAGCAGCCTGTGGGC 360
   ||||| ||||| ||||| ||||| ||||| |||||
Db 155 ArgGlyValGlnIleValMetValIleValThrAspGlyGlnSerHisAsp-----Asn 172
Qy 361 CCCCCATGCAGAGACTC-----AAGACCTGGGCGTCACCGTGTTCATTGTC 408
   ::::: ||||| ||||| ||||| ||||| |||||
Db 173 TyrArgLeuGlnGlnValIleAspIleAspGlyAspGluAsnIleGlnArgPheAlaIle 192
Qy 409 -----AGACCGGCGGAGGCAACTTCCTG-----GAGCTG 438
   ||||| ||||| ||||| ||||| ||||| |||||
Db 193 AlaIleLeuGlySerTySerArgGlyAsnLeuSerThrGluIlePheValGlnGluIle 212
Qy 439 TCAGCGCTGCTTCAGCCCTGCGCAGAGACGCTGACCTTGTGAGCTGGATGAC--- 495
   ::::: ||||| ||||| ||||| ||||| ||||| |||||
Db 213 LysSerIleAlaSerIlePheProthrGluIle-----HisPhePheAsnValSerAspGlu 230
Qy 496 -----CTGCACATCATTTGTCCAGAGCTGAGGGGCTCCATTCTGCG 537
   ||||| ||||| ||||| ||||| ||||| |||||
Db 231 LeuAlaLeuValThrIleValGlnAlaLeuGlyGluArgIlePheAla 246

RESULT 13
type XII collagen alpha-1 chain - eastern newt (fragment)
C/Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
A/Accession: I51027
R/Ref: Yang, Y.; Yang, E.V.; Klatz, K.P.; Tassava, R.A.
Dev. Biol. 168, 503-513, 1995
A/Title: Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII colla
A/Reference number: I51027; MUID:95246925; PMID:7729585
A/Accession: I51027
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-929 <MBI>
A/Cross-references: UNIPROT:Q91145, UNIPARC:UPI000012602E, EMBL:U19494, NID:9632647, PID
F/155-236/Domain: fibronectin type III repeat homology <3FR>
F/631-795/Domain: von Willebrand factor type A repeat homology <VWA3>

Alignment Scores:
Pred. No.: 5,07e-08 Length: 929
Score: 197.50 Matches: 50
Percent Similarity: 48.6% Conservative: 36
Best Local Similarity: 28.2% Mismatches: 90
Query Match: 18.8% Indels: 1
DB: 2 Gaps: 1

US-10-699-035a-1 (1-537) x I51027 (1-929)
Qy 4 GACCTGATGTTCTGCTGACAGCTCAGCAGCGTCTCTACTACAGATTCTCCGGGT 63
   ||||| ::::: ||||| ||||| ||||| |||||
Db 633 AspIleValIleuValAspIleSerTrpSerIleGlyArgProAsnPheIleVal 652
Qy 64 CGGAGTTTGTGGGAGCTGTGCTCCATGCGCTGGGACCGGGGCGCTGCGTGC 123
   ||||| ||||| ||||| ||||| ||||| |||||
Db 653 ArgAsnPheIleSerArgValAlaGlnValPheAspIleGlySerAspArgValGlnIle 672
Qy 124 AGCTGTCGACAGTCGTGGCAGTGCAGCATCAGGAGTTCCTTGGCCAGCAGCAGCTCG 183
   ::::: ||||| ||||| ||||| ||||| |||||
Db 673 AlaAlaSerGlnTySerGlyAspProArgThrGluTrpGlnLeuAsnThrHisTySerThr 692
```

```
Qy 184 GGTGAGCTGCCAGATGCGGTGCTGCTTTCGCCAGCGCATGGTGAACCCACACT 243
   ::::: ||||| ||||| ||||| ::::: |||||
Db 693 LysIleSerLeuMetAspAlaValAlaAsnLeuProTyrlGlyGlyAsnThrHisAsnThr 712
Qy 244 GCGCTGCGCTGTCTATGCCCAAGAACAGCTGTTTGTGAAGATCAGTGGCCCGGCA 303
   ||||| ||||| ||||| ||||| ||||| |||||
Db 713 GlySerAlaLeuLysPheIleLeuGlnAsnAsnAsnHeArgProGlyValGlyMetArgGlu 732
Qy 304 GGGGTGCCCAAGTGTGTGTGAGTGCAGATGGCGGCTCCAGCAGCCTGTGGCCCGC 363
   ||||| ::::: ||||| ||||| ||||| |||||
Db 733 LysAlaArgLysIleAlaIleLeuLeuThrAspGlyLysSerGlnAspAspIleValAla 752
Qy 364 CCATGACAGAGCTCAAGAGACTGGGCGTCAACCGTTCATTGTTCAGCAGCCGCGAGGC 423
   ::::: ||||| ||||| ||||| ||||| |||||
Db 753 ProSerLysArgTyrlAlaAspGluGlyIleGluLeuTyrlAlaValGlyIleLysAsnAla 772
Qy 424 AACTTCTGAGAGCTGTACAGCCGCTGCTCAGCCCTCGCGGAAGACCTGACTTTGTG 483
   ::::: ||||| ||||| ||||| ||||| |||||
Db 773 AspGluAsnGlnIleuLysGlnIleAlaSerAspProAspGlnLeuTyrlMetTyrlAsnVal 792
Qy 484 ---GACCTGATGACCTGACATCATTTGTCCAAAGAGCTGAGGGGCTCATTT 531
   ||||| ||||| ||||| ||||| ||||| |||||
Db 793 AlaAspPheSerLeuLeuThrAsnIleValAsnAspLeuThrGluAsnVal 809

RESULT 14
A45226
Integrin alpha-1 chain - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
A/Accession: A45226
R/Biesecker, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A/Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A/Reference number: A45226; MUID:93155124; PMID:8428973
A/Accession: A45226
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-1151 <BRI>
A/Cross-references: UNIPROT:P56199, UNIPARC:UPI000012D9EB
A/Experimental source: hepatoblastoma cell line HepG2
A/Note: sequence extracted from NCBI backbone (NCBIRef:124326)
F/142-317/Domain: von Willebrand factor type A repeat homology <VWA1>

Alignment Scores:
Pred. No.: 5,49e-08 Length: 1151
Score: 197.00 Matches: 65
Percent Similarity: 47.9% Conservative: 28
Best Local Similarity: 33.5% Mismatches: 81
Query Match: 18.8% Indels: 20
DB: 2 Gaps: 7

US-10-699-035a-1 (1-537) x A45226 (1-1151)
Qy 4 GACCTGATGTTCTGCTGACAGCTCAGCAGCGTCTCTACTACAGATTCTCCGGGT 63
   ||||| ::::: ||||| ||||| ||||| |||||
Db 144 AspIleValIleValIleuAspGlySerAsnSerIle-----TyrProTrpAspSerVal 161
Qy 64 CGGAGTTTGTGGGAGCTGTGCTCCATGCGCTGGGACCGGGGCGCTGCGTGC 123
   ||||| ||||| ||||| ||||| ||||| |||||
Db 162 ThrAlaPheLeuAsnAspLeuLeuLysArgMetAspIleGlyProLysGlnThrGlnVal 181
Qy 124 AGCTGTCGACAGTCGTGGCAGTCGCGCAGCATCAGGAGTTCCTTGGCCAGCAGCAGCTCG 183
   ::::: ||||| ||||| ||||| ||||| |||||
Db 182 GlyIleValGlnTyrglyGlnAsnValThrHisGlnPheAsnLeuAsnLysTySerSer 201
Qy 184 GGTGAGCTGCCAGAGTGTGCTGCTTTCGCCAGCGCATGGT--GACACCCAC 240
   ||||| ||||| ||||| ||||| ||||| |||||
Db 202 ThrGluGlnValIleuValAlaAlaLysLysIleValGlnArgGlyGlyArgGlnThrMet 221
Qy 241 ACTGCGCTGCGCGCTGTATGCAAGAGAACAGCTGTTTGTGAAGATCAGTGGCCCGG 300
   ::::: ||||| ||||| ||||| ||||| |||||
Db 222 ThrAlaLeuGlyIleThrAspThrAlaArgLysGlnAlaPheThrGlnAlaArgIleAlaArg 241
```


THIS PAGE BLANK (USPTO)

RN	[2]

RC TISSUE=Placenta;
 RA Strausberg R.;
 RN Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 RP NUCLEOTIDE SEQUENCE.
 RA Hall R.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC059409; AAH59409.1; -, mRNA.
 DR EMBL; AL391244; CA122657.1; -, Genomic DNA.
 DR EMBL; ENSG00000179403; Homo sapiens.
 DR InterPro: IPR003961; FN III.
 DR InterPro: IPR002035; VWF_A.
 DR Pfam; PF00041; FN3; 2.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS50853; FN3; 2.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 445 AA; 46804 MW; D9FBDE9C0A4DEBDF CRC64;

Alignment Scores:

Pred. No.:	1 03e-59	Length:	445
Score:	902.00	Matches:	178
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	86.0%	Indels:	0
DB:	2	Gaps:	0

US-10-699-035A-1 (1-537) x Q6RCB0_HUMAN (1-445)

OY 1 GGGAGCCTGATGTTCTGCTGAGACAGCTCAGCCAGCTCTCTACTAGAGTTCTCCCGG 60
 DB 33 G1YAplEmetPheuleuaspSerSerAlaSerValSerH1eTgt1uPheserArg 52
 OY 61 GTTCGGAGTTTGTGGGAGAGCTGGGTTCCAGTCCCGGCGGAGCGGCGCTGGCT 120
 DB 53 VALArgGluPhaValG1yGlnLeuValAlaProLeuProLeuG1yThrt1yAlaLeuArg 72
 OY 121 GCCAGTCTGTGACAGCTGGGAGCTGGCCATACACCGAGTTCCCTTCCGAGCAGAC 180
 DB 73 AlAserLeuValH1sValG1ySerArgProTyrtThGluPherProhe1yGlnH1sSer 92
 OY 181 TCGGGTGAAGCTGCCAGGAGATGCGGTGCTCTCTGCCAGCGCATGGGTGACACCCAC 240
 DB 93 SerG1yGluAlaAlaGlnAspAlaValAlaGAlaSerAlaGlnArgMetG1yAspThrH1s 112
 OY 241 ACTGGCTGGCGCTGGTCTATGCGCAAGAAAGCTTTGCTGAAGCATCAGGTGCCCGG 300
 DB 113 ThrG1yLeuAlaLeuValTyrtAlaTySg1uGlnLeuPha1aGlnAlaSerG1yAlaArg 132
 OY 301 CCAGGGGTGCCCAAAGTGTGCTGGTGGTGAAGATGGGCGCTCCAGCAGCCCTGGAGC 360
 DB 133 ProG1yValProTyrtAlaLeuValAlaTyrtAlaH1sAspG1yG1ySerSerAspProValG1y 152
 OY 361 CCCCCCATGAGAGCTCAAGAGCTGGGCGTCAACCGTTCATTTGTCAGACCGGCCGA 420
 DB 153 ProprometGlnG1uLeuTyrtAspLeuG1yValThrtValPhel1eValSerThrtG1yArg 172
 OY 421 GGCAATTCCTGAGAGCTGTCAAGCGCTGCTAGCCCGCTCCGGAAGAAGACTGCACTTT 480
 DB 173 G1yAspPheLeuG1uLeuSerAlaAlaAlaSerAlaProAlaG1uTyrtH1sPheH1sPhe 192
 OY 481 GTGAGCAGTGAAGCTGCACATCTTGTCCAAGAGCTGAGGGGCTCCATTCTC 534
 DB 193 ValAspValAspAspLeuH1s1le1eValGlnG1uLeuH1yG1ySer1leLeu 210

RESULT 2
 Q8R2Z5 MOUSE PRELIMINARY; PRT; 415 AA.
 AC Q8R2Z5
 DT 01-JUN-2002 (TRENBLREL. 21, Created)
 DT 01-JUN-2002 (TRENBLREL. 21, Last sequence update)

DT 13-SEP-2005 (TRENBLREL. 31, Last annotation update)
 DE Von Willebrand factor A domain-related protein (Mus musculus 11 days
 DE pregnant adult female ovary and uterus cDNA, RIKEN full-length
 DE enriched library, clone:5031410123 product:VON WILLEBRAND FACTOR A-
 DE RELATED PROTEIN homolog).
 CN Name=Wval;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Czech II, and FVB/N;
 RC TISSUE=Colon, and
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.N., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Hellon E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalski D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Czech II;
 RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
 RG NIH MGC Project;
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44 (1999).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi U., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kanakawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Macnuc Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Meshima J., Mazzarelli U., Sakamoto N.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";

RL	Nature 409:685-690(2001).	[5]
RN		
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=C57BL/6J; TISSUE=Ovary and uterus;	
RA	the FANTOM Consortium,	
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;	
RT	"Analysis of the mouse transcriptome based on functional annotation of	
RT	60,770 full-length cDNAs."	
RL	Nature 420:563-573(2002).	[6]
RN		
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=C57BL/6J; TISSUE=Ovary and uterus;	
RX	MEBLINE=00493374; Pubmed=11042159; DOI=10.1101/gr.145100;	
RA	Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,	
RA	Komno H., Akiyama J., Nishi K., Kikunai T., Tashiro H., Itoh M.,	
RA	Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,	
RA	Yamanoto K., Matsunoto H., Sakaguchi S., Ikegami T., Kasaiwagi K.,	
RA	Fujiwaka S., Inoue K., Togawa K., Izawa M., Ohara E., Watabiki M.,	
RA	Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,	
RA	Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;	
RT	"RIKEN integrated sequence analysis (RISA) system-384-format	
RT	sequencing pipeline with 384 multiplexed sequencer."	
RL	Genome Res. 10:1175-1177(2000).	[8]
RN		
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=C57BL/6J; TISSUE=Ovary and uterus;	
RA	Adachi J., Aizawa K., Akiyama T., Arakawa T., Bono H., Carninci P.,	
RA	Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,	
RA	Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirokawa T.,	
RA	Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,	
RA	Katoh H., Kawai J., Kojima Y., Kondo S., Komno H., Kouda M., Koya S.,	
RA	Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,	
RA	Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,	
RA	Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,	
RA	Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,	
RA	Togawa A., Takahashi F., Takaku-Akahita S., Takeda Y., Tanaka T.,	
RA	Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;	
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.	[9]
RN		
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=FVB/N; TISSUE=Colon;	
RA	Director MGC Project;	
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: BC026919; AAR26919.1; -; mRNA.	
DR	EMBL: AK077240; BAC36703.1; -; mRNA.	
DR	EMBL: BC036166; AAH36166.1; -; mRNA.	
DR	HSP: P18614; IMHP.	
DR	Ensembl: ENSMUSG00000042116; Mus musculus.	
DR	MGI: MGI:2179729; Wval.	
DR	GO: GO:0005615; C:extracellular space; TAS.	
DR	InterPro: IPR003961; FN_III.	
DR	InterPro: IPR002035; VWF_A.	
DR	Pfam: PF00041; FN3_2.	
DR	Pfam: PF00092; VWF_1.	
DR	PRINTS: PRO0453; VWFADOMAIN.	
DR	SMART: SM00060; FN3_2.	
DR	SMART: SM00327; VWF_1.	
DR	PROSITE: PSS0865; FN3_2.	
DR	PROSITE: PSS0234; VWF_1.	
SO	SEQUENCE 415 AA; 44709 MW; C3887963B2E334FE CRC64;	

```

Best Similarity: 91.5% Conservative: 12
Percent local Similarity: 84.7% Mismatches: 15
Query_Match: 73.0% Indels: 0
DB: 2 Gaps: 0
US-10-699-035A-1 (1-537) x Q8R2Z5_MOUSE (1-415)

QY 1 GGGAGCCTGATGTCCTGCTGGACAGCGTCAGCGAGCGCTCTCACTACGAGTTCTCCGG 60
DB 33 G1YAspLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSerArg 52
QY 61 GTTCGGAGATTGTGTGGGAGAGCTGGTGGCTCCACTGACCGCCCTGGAGCACCGGGGCGCTCGT 120
DB 53 ValArgGluPheValGlyGluLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 72
QY 121 GCCAGTCTGGTGCACGTGGGACGTGGCGCATACACCGAGTTCCCTTCGGCCAGCACAGC 180
DB 73 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 92
QY 181 TCGGCTGAGGCTGCCAGGATCGGCTGCTTGTGCCACGCGCATGGCTGACACCCAC 240
DB 93 SerGlyAlaAlaIleGluAspAlaIleArgValAlaProGlnAspMetGlyAspThrAsn 112
QY 241 ACTGGCGCTGGCGGTGCTCTATGTCACGAAGAACGCTGTTGGTGAAGCATCAGTGCCCG 300
DB 113 ThrGlyLeuAlaLeuAlaTyrAlaArgGluGluLeuPheAlaGluGluAlaGlyAlaArg 133
QY 301 CCAGGGGATGCCCAAGTGTGGTGTGGGTGACAGATGGCGGCTCCAGGACCTGTGGGC 360
DB 133 ProGlyValProIleValLeuValThrValThrAspGlyGlySerSerAspProValGly 153
QY 361 CCCCCCATGACGAGAGCTCAAGAGCCTGGGCGCTCACCGCTTCATTGTACAGACCGCCGA 420
DB 153 ProProMetGlnGluLeuLysAspLeuGlyValThrIlePheIleValSerThrGlyArg 173
QY 421 GGCAACTCTCTGAGAGCTGTACACCGCTGCTCAGCGCTCGCGGAGAAAGCACTGCACATT 480
DB 173 GlyAsnLeuLeuGluLeuLeuAlaIleAlaSerAlaProAlaGluLysLeuHisPhe 193
QY 481 GTGAGCGTGGATGACCTGCACATCATTTGCCAAGAGCTGAGGGGCTCCATT 521
DB 193 ValAspValAlaAspAspLeuProIleIleAlaArgGluLeuArgGlySerIle 209

RESULT 3
Q923K3_MOUSE PRELIMINARY; PRT; 415 AA.
ID ID Q923K3_MOUSE PRELIMINARY; PRT; 415 AA.
AC AC Q923K3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Von Willebrand factor A-related protein.
GN Name=Wval; Synonyms=4932416AlaIrk, Warp;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22057805; PubMed=12062410; DOI=10.1016/S0014-5793(02)02579-6;
RA FitzGerald J., Ting S.T., Bateman J.F.;
RT "WARP a new member of the von Willebrand factor A-domain superfamily
RT of extracellular matrix proteins.";
RL FEBS Lett. 517:61-66(2002).
DR EMBL: AY030094; AAK39350.1; -; mRNA.
DR HSP, P18614; 1MHP.
DR Ensemble; ENSMUSG00000042116; Mus musculus.
DR MGI; MGI:2179729; 4932416AlaIrk.
DR MGI; MGI:2179729; Wval.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00041; fn3; 2.

```


RESULT 6	Q4SCD1_TETNG
ID	Q4SCD1_TETNG PRELIMINARY; PRT; 505 AA.
AC	Q4SCD1;
DT	13-SEP-2005 (TrEMBLrel. 31, Created)
DT	13-SEP-2005 (TrEMBLrel. 31, last sequence update)
DT	13-SEP-2005 (TrEMBLrel. 31, last annotation update)
DE	Chromosome undetermined SCAR14659, whole genome shotgun sequence.
GN	ORFNames=GSTNG00020556001;
OS	Tetradodon nigroviridis (green puffer).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Neuteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neuteleostei;
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC	Tetraodontidae; Tetraodontidae; Tetradodon.
NCBI_TaxID=99883;	

RP NUCLEOTIDE SEQUENCE.
RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicola S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Desilla C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Arnaud V., Ubini Z., Castell C., Castelli V., Karinka M., Vacherie B.,
RA Blemont C., Skalli J.C., Cattolico L., Poulain J., de Bernardis V.,
RA Crnaud C., Duprat S., Broctier P., Coutanceau J.P., Guzy J.,
RA Parra G., Lardier G., Chapprie C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Lander W., Schachner V., Queller F., Sautin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissensbach J., Roest Croliiss H.,
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype".
RL Nature 431:946-957(2004).
RN [2]

RP NUCLEOTIDE SEQUENCE.
RG Genoscope, Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

EMBL: CAAB01014659; CAG01701.1, -, Genomic DNA.
SEQUENCE 505 AA; 55316 MW; C724FAJ99E0751FC CRC64;

US-10-699-035A-1 (1-537) x Q4SCD1_TETNG (1-505)

Qy I GGGACCTGATGTCTCTGTCGACAGCTTAGCCAGGCTCTCTCACTAAGAGTTCTCCGG 60

```

RESULT 7
Q4SXE3_TETNG
ID Q4SXE3_TETNG PRELIMINARY; PRT; 2225 AA.
AC Q4SXE3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF12445, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00010895001;
OS Tetradodon nigrovittatus (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetradodon.
OX NCBI_TaxID=99883;

```

RP Nucleotide sequence.
RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bonneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicoud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segreins B.,
RA Dailly C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Arthouard V., Jubin C., Castelloli V., Karinka M., Vacherie B.,
RA Biemont C., Skalli Z., Castelloli L., Poulan J., De Bernardini V.,
RA Ciraudo C., Duprat S., Brotier P., Coucarneau J.P., Gouzy J.,
RA Parra G., Lardier G., Chappie C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nussbaum C., Kahn D., Robinson-Rechavi M.,
RA Lander P., Schachner V., Queller F., Sautin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Croliuss H.,
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RL the early vertebrate proto-karyotype." ,
RL Nature 431:946-957(2004).

RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 -i- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

CC	preliminary data.
DR	EMBL; CAAB01012445; CAF96689.1; -; Genomic_DNA.
DR	InterPro; IPR0081810; Collagen.
DR	InterPro; IPR0033961; FN_III.
DR	InterPro; IPR0031129; laminin G_TSP_N.
DR	InterPro; IPR001220; lectin_1egb.
DR	InterPro; IPR002035; WMF_A.
DR	Pfam; PF01391; Collagen; 2.
DR	Pfam; PF00041; fn3; 8.
DR	Pfam; PF00092; VWA; 3.
DR	PRINTS; PR00453; WMFDOMAIN.
DR	SMART; SM00060; FN3; 9.
DR	SMART; SM00210; TSPN; 1.
DR	SMART; SM00327; VWA; 3.
DR	PROSITE; PSS0853; FN3; 9.
DR	PROSITE; PSS0307; LECTIN_LEGUME_BETA; UNKNOWN_1.
DR	PROSITE; PSS0234; VWA; 3.
FT	Collagen; Extracellular matrix; structural protein.
FT	NON_TER
SQ	SEQUENCE 2225 AA; 239674 MW; EC6545B60A778BD9 CRC64

QY 184 GGTAGAGCTCCAGAGATGCGTGTCTTGTCCCGCAGGAGTGGTGAACCCACACT 243
 Db 498 LeuGluValValThraValLeuValArgThrPheProTyrArgGlyGlySerThrAsnThr 517
 QY 244 GGGCGGGCGGTGTATGATCCAGAGCGTGTGCTGAAGCATCATGAGTCCCGGCCA 303
 Db 518 GlyArgAlaMetThrTyrValArgGluThrValPheGlnAlaSerArgGlyAlaArgAla 537
 QY 304 GGGGCGCCCAAGTGTGTGTGTGTGTGACAGATGGCGGCTCCAGCGACCTGTGGGCCCC 363
 Db 538 HisValProArgValThrIleuLeuIleThrAspGlyLysSerSerAspAlaPheGlnGlu 557
 QY 364 CCCATGACGAGACTCAAGAGCTGGGCGGTACCGCTGTTTATTGTACAGACCGCGGAGGC 423
 Db 558 ProAlaAlaAsnLeuArgAsnSerAspValGluIlePheAlaValGlyValLysAspAla 577
 QY 424 AACTTCCTGAGAGCTGTGACCGCTGCGTCCAGCGCTCCAGAGACCTGTGACCTTTGG 483
 Db 578 ValArgSerGluLeuGluAlaIleAlaAsnAlaProAlaGluThrHisValTyrThrVal 597
 QY 484 --GACGTGATGACCTGCACATCTTGTCCAAAGAGTGAAGGCGCTCAT 531
 Db 598 GluAspPheAspAlaPheGlnArgIleSerThrGluLeuThrGlnSerIle 614

RESULT 9

COEAL CHICK STANDARD; PRT; 1888 AA.

AC P32018; OELBLD;
 DT 01-JUN-1993 (Rel. 26, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)

DE Collagen alpha 1(XIV) chain precursor (Undulin).
 GN Name=COL14A1;

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.

OX NCBI_TaxID=9031;
 RN (1)

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Embryo;
 RA MEDLINE=93185668; PubMed=8444186; Winterhalter K.H., Trueb B.,
 RT Waechli C., Trueb J., Kessler B.,
 RL Eur. J. Biochem. 212:483-490(1993).

[2]
 RN NUCLEOTIDE SEQUENCE OF 402-1549,
 RP MEDLINE=92339443; PubMed=1339349;
 RA Trueb J., Trueb B.,
 RL Eur. J. Biochem. 207:549-557(1992).

[3]
 RN NUCLEOTIDE SEQUENCE OF 1582-1770,
 RP MEDLINE=92037585; PubMed=1935930;
 RA Gordon M., Castagnola P., Dublet B., Linsemayer T.F.,
 RT van der Rest M., Mayne R., Olsen B.R.,
 RT "Cloning of the cDNA for a new member of the class of fibril-
 associated collagens with interrupted triple helices";
 RL Eur. J. Biochem. 201:333-338(1991).

[4]
 RN NUCLEOTIDE SEQUENCE OF 1582-1770,
 RP Apte S.S.,
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.

[5]
 RP STRUCTURE BY NMR OF 1853-1885,
 RX MEDLINE=98357967; PubMed=9694594; DOI=10.1016/S0945-053X(98)90027-0;
 RA Gly-Lozinger C., Aubert-Foucher E., Penin F., Deleage G., Dublet B.,
 RT "Identification and characterization of a heparin binding site within
 the NCI domain of chicken collagen XIV";
 RL Matrix Biol. 17:145-149(1998).

[6]
 RP STRUCTURE BY NMR OF 1852-1885.

RX MEDLINE=99280705; PubMed=10350466; DOI=10.1021/bi9900222;
 RA Montserret R., Aubert-Foucher E., McLeish M.J., Hill J.M., Fichoux D.,
 RA Jaouen M., van der Rest M., Deleage G., Penin F.;
 RT "Structural analysis of the heparin-binding site of the NCI domain of
 collagen XIV by CD and NMR";
 RL Biochemistry 38:6479-6488(1999).
 CC -1- FUNCTION: An adhesive role by integrating collagen bundles. It is
 CC probably associated with the surface of interstitial collagen
 CC fibrils via COL1. The COL2 domain may then serve as a rigid arm
 CC which sticks out from the fibril and protrudes the large N-
 CC terminal globular domain into the extracellular space, where it
 CC might interact with other matrix molecules or cell surface
 CC receptors.
 CC -1- SUBUNIT: Homotrimer (Probable).
 CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -1- TISSUE SPECIFICITY: Wide tissue distribution; high presence in
 CC dense connective tissue in skeletal muscle.
 CC -1- PM: Lysines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
 CC -1- PM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- PTM: May contain numerous cysteine residues involved in inter- and
 CC intramolecular disulfide bonding.
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with
 CC interrupted helices (FACIT) family.
 CC -1- SIMILARITY: Contains 8 fibronectin type-III domains.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -1- SIMILARITY: Contains 2 VWFA domains.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.

CC -----
 CC EMBL: X70793; CAA50064.1; -; mRNA.
 CC EMBL: X70792; CAA50063.1; -; mRNA.
 CC EMBL: X66138; CAA46928.2; -; mRNA.
 CC EMBL: X65122; CAA46238.1; -; mRNA.
 CC PIR: A45974; A45974.
 CC PIR: S78476; S78476.
 CC PDB: 1B9P; NMR; A=1853-1885.
 CC PDB: 1B9Q; NMR; A=1853-1885.
 CC InterPro: IPR008160; Collagen.
 CC InterPro: IPR003961; FN_III.
 CC InterPro: IPR002035; VWF_A.
 CC Pfam: PF01391; Collagen; 4.
 CC Pfam: PF00041; fn3; 8.
 CC Pfam: PF00092; VWF; 2.
 CC PRINTS: PR00453; VWFADOMAIN.
 CC PROSITE: PS50853; FN3; 8.
 CC PROSITE: PS50234; VWFA; 2.
 CC 3D-structure: Cell adhesion;
 CC Glycoprotein; Hydroxylation;
 CC Repeat; Signal; Structural protein.
 CC Potential.
 CC Collagen alpha 1(XIV) chain.
 CC Fibronectin type-III 1.
 CC Fibronectin type-III 1.
 CC Fibronectin type-III 2.
 CC Fibronectin type-III 3.
 CC Fibronectin type-III 4.
 CC Fibronectin type-III 5.
 CC Fibronectin type-III 6.
 CC Fibronectin type-III 7.
 CC Fibronectin type-III 8.
 CC Fibronectin type-III 9.
 CC VWFA 2.
 CC TSP N-terminal.
 CC Nonhelical region (NC4).
 CC Triple-helical region 1 (COL2).
 CC Triple-helical region 2 (COL1).
 CC Cell attachment site (Potential).
 CC Cell attachment site (Potential).
 CC Poly-Thr.
 CC COMBIA5 721 733

RESULT 11	MATN4_MOUSE	STANDARD;	PRT;	624 AA.
ID	MATN4_MOUSE	089029; Q9QWS3;		
AC	089029; 089030; Q9QWS3;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	10-MAY-2005 (Rel. 47, Last annotation update)			
DE	Matrilin-4 precursor (MAT-4).			
GN	Name=Matn4;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;			
OC	Muroidea; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RN	NUCLEOTIDE SEQUENCE (ISOFORMS LONG AND SHORT), AND VARIANTS.			
RC	STRAIN=C57BL/6J, and CD-1; TISSUE=Fetal;			
RX	MEDLINE=96442849; PubMed=9771906; DOI=10.1016/S0014-5793(98)01111-9;			
RA	Wagner R., Kobbe B., Paulsson M.;			
RT	"Matrilin-4, a new member of the matrilin family of extracellular			
RT	matrix proteins.";			
RL	FEBS Lett. 436:123-127(1998).			
RN	[2]			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM LONG).			
RC	STRAIN=C57BL/6J; TISSUE=Mammary gland;			
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RA	Stauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Aleisch S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diachenko L., Marinina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Donald M.F., Cabavant T.L., Schetz T.E.,			
RA	Brownstein M., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Boeak S.A., McGown P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,			
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Falley U., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield V.S.N., Krzywinski M.T., Skalska U., Smallus D.E.,			
RA	Schmerch A., Schin J.E., Jones S.J.W., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[3]			
RP	NUCLEOTIDE SEQUENCE OF 218-257.			
RX	MEDLINE=99043341; PubMed=9827539; DOI=10.1016/S0014-5793(98)01293-9;			
RA	Wagner R., Kobbe B., Paulsson M.;			
RT	"Genomic organisation, alternative splicing and primary structure of			
RT	human matrilin-4.";			
RL	FEBS Lett. 438:165-170(1998).			
CC	-1- FUNCTION: Major component of the extracellular matrix of			
CC	cartilage.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Name=Long;			
CC	Isoid=089029-1; Sequence=Displayed;			
CC	Name=Short;			
CC	Isoid=089029-2; Sequence=VSP 001401;			
CC	-1- TISSUE SPECIFICITY: Lung, brain, sternum, kidney and heart.			
CC	-1- DEVELOPMENTAL STAGE: The short isoform was detected in 7 weeks old			
CC	mice but not in developing mice (19.5 dpc embryos or in 2, 8, and			
CC	21 days old animals).			
CC	-1- SIMILARITY: Contains 4 EGF-like domains.			
CC	-1- SIMILARITY: Contains 2 VWFA domains.			
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use as long as its content is in no way modified and this statement is not			

[illegible]

```

Db      76 GlyValIleGlnTyrSerSerGlnValGlnSerValPheProLeuGlyAlaPheSerArg 95
Oy      184 GGTGAGGCTGCCAGATGCGGTGCTGCTTGTGCCAGCCGATGGGTGACACCACT 243
Db      96 ArgGluAspMetGlnArgAlaIleArgAlaValProLeuAlaGlnGlyThreThr 115
Oy      244 GGCCTGGCGCTGCTGATCCAGAGACAGCTGTTGTCGAAGCATCAGTCCCGGCCA 303
Db      116 GlyLeuAlaIleGlnTyrAlaIleAsnValAlaPheSerGluAlaGlnGlyAlaArgPro 135
Oy      304 GGG-----GTGCCCAAGTCTGCTGTGGGTGACAGATGGCGGCTCCAGCACTT 354
Db      136 SerGluGluArgValProArgValLeuValIleValThrAspGlyArgProGlnAspArg 155
Oy      355 GTGGGCCCCCGCATGACAGAGCTCAAGAGACCTGGCGGCTGTCATGTCACCACT 414
Db      156 ValAlaGluValAlaAlaGlnAlaArgAlaArgGlyIleGluIleTyrAlaValGlyVal 175
Oy      415 GGCCTGAGGCAACTTCTGAGCTGTGACGCCGCTGCTCAGCCCTGCGAGAGACCTGT 474
Db      176 GlnArgAlaAspValGlySerLeuArgThrMetAlaSerProProLeuAspGlnHisVal 195
Oy      475 CACTTGTGAGACGTGATGATGACATCATTTGTCGAAG 516
Db      196 PheLeuValGluSerPheAspLeu-----IleGlnGlu 206

RESULT 12
COCAL CHICK STANDARD; PRT; 3124 AA.
ID      COCAL CHICK 013944; 004509;
AC      01-JAN-1990 (Rel. 13, Created)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      10-MAY-2005 (Rel. 47, Last annotation update)
GN      Collagen alpha 1(XII) chain precursor (Fibrochimerin).
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
CX      NCBI_TaxId=9031;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=White Leghorn;
RX      MEDLINE=92011862; PubMed=1918137; DOI=10.1083/jcb.115.1.209;
RA      Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
RA      Nishida Y., Obara M., Kimata K.;
RT      "The complete primary structure of type XII collagen shows a chimeric
RT      molecule with reiterated fibronectin type III motifs, von Willebrand
RT      factor A motifs, a domain homologous to a noncollagenous region of
RT      type IX collagen, and short collagenous domains with an Arg-Gly-Asp
RT      site."
RT      J. Cell Biol. 115:209-221 (1991).
RN      [2]
RP      NUCLEOTIDE SEQUENCE OF 2456-3124, AND PROTEIN SEQUENCE OF 2772-2794
RP      AND 2846-2873.
RX      MEDLINE=90062079; PubMed=2584192;
RA      Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
RT      "Type XII collagen. A large multidomain molecule with partial homology
RT      to type IX collagen."
RT      J. Biol. Chem. 264:19772-19778 (1989).
RN      [3]
RP      NUCLEOTIDE SEQUENCE OF 2960-3076.
RX      MEDLINE=87317590; PubMed=3476925;
RA      Gordon M.K., Gerecke D.R., Olsen B.R.;
RT      "Type XII collagen: distinct extracellular matrix component discovered
RT      by cDNA cloning."
RT      Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044 (1987).
RN      [4]
RP      NUCLEOTIDE SEQUENCE OF 1-1283 (ISOFORM SHORT), AND ALTERNATIVE
RP      SPLICING.
RC      TISSUE=Embryo;
RX      MEDLINE=93042014; PubMed=1420368; DOI=10.1016/0167-4781(92)90145-P;

```

```

RA      Trueb J., Trueb B.;
RT      "The two splice variants of collagen XII share a common 5' end.";
RT      Biochim. Biophys. Acta 1171:97-98 (1992).
RN      [5]
RP      ALTERNATIVE SPLICING.
RX      MEDLINE=95370352; PubMed=7642694; DOI=10.1083/jcb.130.4.1005;
RA      Koch M., Bohmann B., Mathison M., Hagios C., Trueb B., Chiquet M.;
RT      "Large and small splice variants of collagen XII: differential
RT      expression and ligand binding."
RT      J. Cell Biol. 130:1005-1014 (1995).
CC      -1- FUNCTION: Type XII collagen interacts with type I collagen-
CC      containing fibrils, the COL1 domain could be associated with the
CC      surface of the fibrils, and the COL2 and NC3 domains may be
CC      localized in the perifibrillar matrix.
CC      -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC      nontriple-helical sequences.
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Comment=The final tissue form of collagen XII may contain
CC      homotrimers of either isoform long or isoform short or any
CC      combination of isoform long and isoform short. Only isoform long
CC      is a proteoglycan. Isoform long has more restricted expression
CC      in embryonic tissue than isoform short;
CC      Name=Long;
CC      IsoId=PI3944-1; Sequence=Displayed;
CC      Name=Short;
CC      IsoId=PI3944-2; Sequence=VSP 001148;
CC      -1- TISSUE SPECIFICITY: Type XII collagen is present in tendons,
CC      ligaments, perichondrium, and perosteum, all dense connective
CC      tissues containing type I collagen.
CC      -1- DOMAIN: This sequence defines five distinct domains, two triple-
CC      helical domains (COL1 and COL2) and three nontriple-helical
CC      domains (NC1, NC2, and NC3).
CC      -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC      each end.
CC      -1- PTM: Prolines at the third position of the tripeptide repeating
CC      unit (G-X-Y) are hydroxylated in some or all of the chains.
CC      -1- PTM: O-glycosylated; glycosaminoglycan of chondroitin-sulfate type
CC      (By similarity).
CC      -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC      interrupted helices (FACIT) family.
CC      -1- SIMILARITY: Contains 18 fibronectin type-III domains.
CC      -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC      -1- SIMILARITY: Contains 4 WFPA domains.
CC      -----
CC      This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use as long as its content is in no way modified and this statement is not
CC      removed.
CC      -----
DR      EMBL, D00824; BAA00701.1; -; mRNA.
DR      EMBL, X61024; CAA43358.1; -; mRNA.
DR      EMBL, M17375; AAA48718.1; -; mRNA.
DR      EMBL, J05137; AAA48635.1; -; mRNA.
DR      EMBL, X67327; CAA47744.1; -; mRNA.
DR      PIR, A40020; A40020.
DR      HSSP, P56199; 10CS.
DR      Ensemble; ENSGALG00000015908; Gallus gallus.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR003961; FN_III.
DR      InterPro; IPR003129; Laminin_G_TSP_N.
DR      InterPro; IPR002035; VWF_A.
DR      Pfam, PF01391; Collagen; 4.
DR      Pfam, PF00041; Fn3; 17.
DR      Pfam, PF00092; VWA; 4.
DR      PRINTS, PR00453; VWFADOMAIN.
DR      SMART, SMO0060; FN3; 18.
DR      SMART, SMO0210; TSPN; 1.
DR      SMART, SMO0327; VWA; 4.
DR      PROSITE, PS50853; FN3; 18.
DR      PROSITE, PS50234; VWFA; 4.
KW      Alternative splicing; Cell adhesion; Collagen;

```

KW Direct protein sequencing; Extracellular matrix; Glycoprotein;
 KM Hydroxylation; Repeat; Signal; Structural protein.
 FT SIGNAL 1 24 Potential.
 FT CHAIN 25 3124 Collagen alpha 1(XII) chain.
 FT DOMAIN 25 112 Fibronection type-III 1.
 FT DOMAIN 139 311 VFPA 1.
 FT DOMAIN 332 421 Fibronection type-III 2.
 FT DOMAIN 439 615 VFPA 2.
 FT DOMAIN 630 718 Fibronection type-III 3.
 FT DOMAIN 721 809 Fibronection type-III 4.
 FT DOMAIN 812 902 Fibronection type-III 5.
 FT DOMAIN 905 993 Fibronection type-III 6.
 FT DOMAIN 995 1083 Fibronection type-III 7.
 FT DOMAIN 1086 1175 Fibronection type-III 8.
 FT DOMAIN 1199 1371 VFPA 3.
 FT DOMAIN 1386 1472 Fibronection type-III 9.
 FT DOMAIN 1474 1564 Fibronection type-III 10.
 FT DOMAIN 1566 1654 Fibronection type-III 11.
 FT DOMAIN 1655 1745 Fibronection type-III 12.
 FT DOMAIN 1756 1845 Fibronection type-III 13.
 FT DOMAIN 1847 1935 Fibronection type-III 14.
 FT DOMAIN 1937 2026 Fibronection type-III 15.
 FT DOMAIN 2028 2117 Fibronection type-III 16.
 FT DOMAIN 2119 2206 Fibronection type-III 17.
 FT DOMAIN 2210 2294 Fibronection type-III 18.
 FT DOMAIN 2327 2500 VFPA 4.
 FT DOMAIN 2524 2716 TSP N-terminal.
 FT DOMAIN 2751 2902 Nonhelical region (NC3).
 FT REGION 2903 2945 Triple-helical region (COL2) with 1
 FT REGION 2946 3048 Triple-helical region (COL1) with 2
 FT REGION 3049 3124 Nonhelical region (NC1).
 FT MOTIF 2899 2901 Cell attachment site (Potential).
 FT COMPBIAS 3086 3096 Imperfections.
 FT COMPBIAS 3111 3123 Nonhelical region.
 FT CARBOHYD 32 32 Arg/Lys-rich (acidic).
 FT CARBOHYD 797 797 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 890 890 O-linked (Xyl...) (chondroitin sulfate) (Potential).
 FT CARBOHYD 890 890 O-linked (Xyl...) (chondroitin sulfate) (Potential).
 FT CARBOHYD 981 981 O-linked (Xyl...) (chondroitin sulfate) (Potential).
 FT CARBOHYD 1006 1006 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1032 1032 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1044 1044 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1512 1512 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1767 1767 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 2210 2210 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 2273 2273 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 2532 2532 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 2683 2683 N-linked (GlcNAc...) (Potential).
 FT VARSPLIC 25 1188 Missing (in isoform short).
 FT CONFLICT 1258 1258 /FTID=VSP_001148.
 FT CONFLICT 1264 1264 T -> S (in Ref. 4).
 FT CONFLICT 2759 2759 D -> E (in Ref. 4).
 FT CONFLICT 2803 2803 P -> A (in Ref. 2).
 FT CONFLICT 2977 2977 L -> F (in Ref. 2).
 FT CONFLICT 3075 3076 V -> F (in Ref. 2).
 FT CONFLICT 3075 3076 QP -> AG (in Ref. 3).
 SO SEQUENCE 3124 AA; 340582 MW; 094285AF67E346CF CRC64;

Alignment Scores:
 Pred. No.: 2,236-11 Length: 3124
 Score: 263.50 Matches: 62
 Percent Similarity: 53.1% Conservative: 32
 Best Local Similarity: 35.0% Mismatches: 82
 Query Match: 25.1% Indels: 1
 DB: 1 Gaps: 1

US-10-699-035A-1 (1-537) x COCA1_CHICK (1-3124)

OY 4 GACCTGATGTTCCCTGCTGAGACGCTCAGCCAGCCTCTCTCACTACAGATTCTCCGGGTT 63
 Db 439 AspvAlValPheLeuValAspGlySerTyrSerIleGlyLeuAlaAspPheValValVal 458
 OY 64 CGGAGGTTTGGGGGCGAGCTGGTGCTCCACTGCCGCCCGGACCGGGGCGCTGAGTCC 123
 Db 459 ArgAlaPheLeuGluValLeuValLysSerPheGluLysSerProAlaGlyValGlnLe 478
 OY 124 AGTGTGTCAGCTGGGCGAGTGGCCATACCCAGATGTCCTTCCTGGCCAGCAGACTCG 183
 Db 479 SerLeuValGlnTyrSerArgAspProIleMetGluPheSerLeuAlaArgTyrAlaArg 498
 OY 184 GGTAGGCTCCCGAGATCCGCTGCTTCTTCCAGCCAGCTGAGTGAACCCACT 243
 Db 499 ValLysAspIleIleGlnAlaIleAlaIleAlaPheProTyrArgGlyGlySerThrAsnThr 518
 OY 244 GGCTGGCGGCTGCTGTATGCCAGGAACAGCTGTTGTCAGATCAGTGGCCGGCA 303
 Db 519 GlyLysAlaMetThrTyrValAlaGluLysValPheValThrSerLysGlySerArgPro 538
 OY 304 GGGTGCCCAAGTGTCTGTGGTGTGACAGATGGCGCTCCAGCAGCCTGTGGGCCCC 363
 Db 539 AsnValProArgValMetIleLeuIleThrAspGlyLysSerSerAlaPheLysGlu 558
 OY 364 CCATGACGAGCTCAAGACCTGGCGCTCAACCTGTTCAATTGCAGCAGCCGCGAGGC 423
 Db 559 ProAlaIleLysLeuArgAspAlaAspValGluIlePheAlaValGlyValLysAspAla 578
 OY 424 AACCTTCGAGCTGTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
 Db 579 ValArgThrGluLeuGluAlaIleAlaIleAlaSerProAlaGluThrAlaValThrVal 598
 OY 484 --GACGTGATGACCTGCACATCATTTGCCAAGAGTGGAGGCTTCATT 531
 Db 599 GluAspPheAspAlaPheGluArgIleSerPheGluLeuThrGlnSerVal 615

RESULT 13
 Q4SH63_TESTNG PRELIMINARY; PRT; 1557 AA.
 ID Q4SH63; AC Q4SH63;
 DT 13-SEP-2005 (TREMBlrel. 31, Created)
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
 DE Chromosome 8 SCAP14587, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG00018311001;
 OS Tetraodon nigroviridis (Green puffer).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 Nicard S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 Dastiv S., Salanoubat M., Levy M., Boudet N., Castellano S.,
 Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 Biemont C., Skalli Z., Carollino L., Poulain J., de Berardinis V.,
 Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosaek S.,
 Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 Landet V., Schachter V., Queller F., Saurin W., Scarpelli C.,
 Winkler P., Lander E.S., Weissenbach J., Roest Crolius H.,
 "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype."
 RL Nature 431:946-957(2004).
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (Feb-2004) to the EMBL/Genbank/DBJ databases.

CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CA601014587; CAG00019.1; -; Genomic_DNA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR00129; Laminin_G_TSP_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00092; VWA; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; VWA; 2.
DR PROSITE; PS50853; FN3; 7.
DR PROSITE; PS50234; VWF; 2.
KM Collagen; Extracellular matrix; Structural protein.
FT NON_TER
SQ SEQUENCE 1557 AA; 168266 MW; 6203E9A0900C6D9D CRC64;

Alignment Scores:
Pred. No.: 2,29e-11 Length: 1557
Score: 263.00 Matches: 61
Percent Similarity: 56.3% Conservative: 37
Best Local Similarity: 35.1% Mismatches: 74
Query Match: 25.1% Indels: 2
DB: Gaps: 2

US-10-699-035a-1 (1-537) x Q4SH63_TETNG (1-1557)

QY 4 GACCTGATGTTCTCTGTGACAGCTGACGCGTCTCTACACGAGTTCCTCCGGATT 63
DB 891 Aspleivalpheleuvalaspolyserilserileglyaspulasmhethelysile 910
QY 64 CGGAGTTGTGGGGGAGCTGTGCTCCACCTGCGCTG--GGACCGGGGCGCTGCGT 120
DB 911 ThrArgpheleuhisetherThralGlyserleuaspheulieglyThraspGlyThrGln 930
QY 121 GCGAGTCTGTGACGCTGGGAGTGGCCATACAGCGATGCCCTTGGCGGACGACG 180
DB 931 ValAlaileAlaGlnPheSeraspaspAlaArgThgluhPheGlnLeuSerSerHisSer 950
QY 181 TCGGATGAGCTGCTCCAGATGCGGTGCTGCTTGTCCCGACGATGGGTGACACCCAC 240
DB 951 AsnLysgluAlaLeuLeuGluAlaileGlnLysIleSerTrlyspelyGlyAsnThrLys 970
QY 241 ACTGCGCTGCGCTGTCTATGCAAGAAACAGCTTTGTCTGAAGCATCAGGTGCGG 300
DB 971 ThrGlyArgAlaileLysHisValLysGlnSerIlePheSerleuGlnIleGlyAlaArg 990
QY 301 CCAGGGGTGCGCAAAATGCTGTGTGTGACAGATGGGCGCTCCGACGACCTGTGGG 360
DB 991 ArgGlyValProLysValLeuValValLeuHisAspGlyArgSerGlnaspValAsn 1010
QY 361 CCCCCATCAGAGCTCAAGACGCTGCGGTGCTGCTTGTCTGACACCGGCGCA 420
DB 1011 LysValSerLysGlnMetGlnMetaspGlyTrIleIlePheAlaileGlyPheAlaasp 1030
QY 421 GGCACCTTCCTGAGCTGTCAAGCGGCTGCTGACCGCTCCGACGACGACCTGACTT 480
DB 1031 AlaAspTrGlyGluLeuValAsnIleAlaSerLysProSeraspArgHisValPhePhe 1050
QY 481 GNG--GAGCTGATGACCTGCAATCATCTGTCACAGACTG 519
DB 1051 ValaspaspLeuaspAlaValLysLysIleGlnGlnGlnLeu 1064

RESULT 14
OSVYK2_HUMAN PRELIMINARY; PRT; 2884 AA.
AC OSVYK2;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Collagen, type XII, alpha 1.
GN Name=COL12A1; ORFNames=RP1-238D15.1-003;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OC NCBI_TaxID=9606;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RA Tromans A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN (2)
RP NUCLEOTIDE SEQUENCE.
RA Smith M.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN (3)
RP NUCLEOTIDE SEQUENCE.
RA Corby N.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL354664; CAH71309.1; -; Genomic_DNA.
DR EMBL; AL080250; CAI19897.1; -; Genomic_DNA.
DR EMBL; AL096771; CAI19907.1; -; Genomic_DNA.
DR EMBL; AL080250; CAH71309.1; JOINED; Genomic_DNA.
DR EMBL; AL354664; CAH71309.1; JOINED; Genomic_DNA.
DR EMBL; AL354664; CAI19897.1; JOINED; Genomic_DNA.
DR EMBL; AL080250; CAI19907.1; JOINED; Genomic_DNA.
DR EMBL; AL354664; CAI19907.1; JOINED; Genomic_DNA.
DR EMBL; AL096771; CAI19897.1; JOINED; Genomic_DNA.
DR Ensemble; ENSG00000111799; Homo sapiens.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0005578; Extracellular matrix (sensu Metazoa); IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00041; fn3; 18.
DR Pfam; PF00092; VWA; 4.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 18.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; VWA; 4.
DR PROSITE; PS50853; FN3; 18.
DR PROSITE; PS50234; VWF; 4.
KM Collagen; Extracellular matrix; Repeat; Structural protein.
SQ SEQUENCE 2884 AA; 315869 MW; 2D598F1365EB454D CRC64;

Alignment Scores:
Pred. No.: 3.14e-11 Length: 2884
Score: 261.50 Matches: 64
Percent Similarity: 52.5% Conservative: 29
Best Local Similarity: 36.2% Mismatches: 83
Query Match: 24.9% Indels: 1
DB: Gaps: 1

US-10-699-035a-1 (1-537) x OSVYK2_HUMAN (1-2884)

QY 4 GACCTGATGTTCTCTGTGACAGCTGACGCGTCTCTACACGAGTTCCTCCGGATT 63
DB 440 AspleivalpheleuvalaspolyserilserileglyIleAlaIasnhethelysVal 459
QY 64 CGGAGTTGTGGGGGAGCTGTGCTCCACCTGCGCTGCGGACGCGGCGCTGCGTGC 123
DB 460 ArgAlapheleuGluValLeuValLysSerPheGlnIleSerProAsnArgValGlnIle 479
QY 124 ACTGTGTCGACAGTGGGACGATGCGCATACAGAGTTCCTTGGCGGACGACGCTG 183
DB 480 SerleuValGlnTrYSerArgaspProHisTrnGlnPheThrLeuLysPheThrLys 499

```

QY 184 GGTAGGCTCCAGATGCGGTGCTTCTGCCAGCCGATGGTGACACCCACT 243
D 500 ValGluSplrleIleGluAlaIleAsnRhpProTyrArgGlySerThrAsnRhp 519
QY 244 GGCGTGGCGGTGCTATGCCAGACAGCGTTTGTGAAAGATCAAGTCCCGGCA 303
D 520 GlyYsAlaMetCmTyrValArgGlySerIlePheValProSerLysGlySerSer 539
QY 304 GGGGTCCCAAAAGTCTGGTGTGACAGATGGCGGCTCCAGCAGCCCTGGGCCCC 363
D 540 AsnValProLysValMetIleLeuIleThrAspGlySerSerAspAlaPheArgAsp 559
QY 364 CCCATGACGAGGCTCAAGACCTGGCGCTCACCGTTTCATTGTAGACAGCCGAGGC 423
D 560 ProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValLysAspAla 579
QY 424 AACTTCCGAGGCTGACCGGCTCCAGCCCTCCGCGGAGAGCACCCTGACTTTGTG 483
D 580 ValArgSerGluLeuGluAlaIleAlaSerProProAlaGluThrAlaPheThrVal 599
QY 484 ---GACGTGATGACCTGCACATCTTGTCCAGAGAGTGGGCTCCATT 531
D 600 GluAspPheAspAlaPheGluArgIleSerPheGluLeuThrGlnSerIle 616

RESULT 15
COCAL HUMAN STANDARD; PRT; 3063 AA.
ID 099715; 099716;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN Name=COL12A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;
OC Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=97288521; PubMed=9143499; DOI=10.1006/geno.1997.4638;
RA Gerecke D.R., Olson F.F., Koch W., Knoll J.H.M., Taylor R.,
RA Hudson D.L., Champilaud M.-F., Olsen B.R., Burgeson R.E.;
RT "Complete primary structure of two splice variants of collagen XII,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(XI)
RT collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human
RT chromosome 6q12-q13."
RL Genomics 41:236-242(1997).
CC -1- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the perifibrillar matrix (By similarity).
CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=The final tissue form of collagen XII may contain
CC homotrimers of either isoform long or isoform short or any
CC combination of isoform long and isoform short;
CC Name=Long;
CC IsoId=Q99715-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q99715-2; Sequence=VSP_001149;
CC -1- TISSUE SPECIFICITY: Found in collagen I-containing tissues: both
CC isoform short and isoform long appear in amnion, chorion, skeletal
CC muscle, small intestine, and in cell culture of dermal
CC fibroblasts, keratinocytes and endothelial cells. Only the short
CC isoform is found in lung, placenta, kidney and a squamous cell
CC carcinoma cell line.
CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC each end (By similarity).
CC -1- PTM: Prolines at the third position of the tripeptide repeating

```

```

CC unit (G-X-Y) are hydroxylated in some or all of the chains (By
CC similarity).
CC -1- PTM: O-glycosylation of isoform long; glycosaminoglycan of
CC chondroitin-sulfate type (By similarity).
CC -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC interrupted helices (FACIT) family.
CC -1- SIMILARITY: Contains 18 fibronectin type-III domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -1- SIMILARITY: Contains 4 VWFA domains.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL: U73778; AAC51244.1; -; mRNA.
DR EMBL: U73779; AAD40483.1; -; mRNA.
DR HSSP: P18614; IMHP.
DR Ensembl: ENSG00000111799; Homo sapiens.
DR HGNC: HGNC:2188; COL12A1.
DR MIM: 120320; -.
DR GO: GO:0005595; C:collagen type XII; TAS.
DR GO: GO:0001501; P:skeletal development; TAS.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR02035; VWF A.
DR Pfam: PF01391; Collagen; 4.
DR Pfam: PF00041; fn3; 18.
DR Pfam: PF00092; VWF; 4.
DR PRINTS: PR00453; VWFADOMAIN.
DR PROSITE: PS50853; FN3; 18.
DR PROSITE: PS50234; VWFA; 4.
DR KW: Alternative splicing; Cell adhesion; Collagen;
KW Direct protein sequencing; Extracellular matrix; Glycoprotein;
KW Hydroxylation; Repeat; Signal; Structural protein.
FT SIGNAL 1 24
FT CHAIN 1 25
FT DOMAIN 25 112
FT DOMAIN 140 316
FT DOMAIN 333 422
FT DOMAIN 440 616
FT DOMAIN 631 719
FT DOMAIN 722 810
FT DOMAIN 813 901
FT DOMAIN 904 993
FT DOMAIN 995 1083
FT DOMAIN 1086 1175
FT DOMAIN 1199 1371
FT DOMAIN 1384 1472
FT DOMAIN 1474 1563
FT DOMAIN 1565 1652
FT DOMAIN 1654 1743
FT DOMAIN 1752 1841
FT DOMAIN 1843 1931
FT DOMAIN 1933 2022
FT DOMAIN 2024 2113
FT DOMAIN 2115 2202
FT DOMAIN 2206 2290
FT DOMAIN 2323 2496
FT DOMAIN 2520 2712
FT DOMAIN 2747 2898
FT REGION 2899 2941
FT REGION 2942 3044
FT REGION 3045 3063
FT MOTIF 862 864
FT MOTIF 2779 2781
FT MOTIF 2895 2897
FT MOD_RES 2944 2944
FT MOD_RES 2947 2947

```

Search completed: February 13, 2006, 13:38:53
Job time : 147.574 secs

```
FT MOD_RES 2950 2950 Hydroxyproline (By similarity).
FT MOD_RES 2959 2959 Hydroxyproline (By similarity).
FT MOD_RES 2965 2965 Hydroxyproline (By similarity).
FT MOD_RES 2968 2968 Hydroxyproline (By similarity).
FT MOD_RES 2971 2971 Hydroxyproline (By similarity).
FT MOD_RES 2983 2983 Hydroxyproline (By similarity).
FT MOD_RES 3000 3000 Hydroxyproline (By similarity).
FT MOD_RES 3003 3003 Hydroxyproline (By similarity).
FT MOD_RES 3014 3014 Hydroxyproline (By similarity).
FT MOD_RES 3023 3023 Hydroxyproline (By similarity).
FT MOD_RES 3026 3026 Hydroxyproline (By similarity).
FT MOD_RES 3029 3029 Hydroxyproline (By similarity).
FT CARBOHYD 700 700 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 798 798 O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT CARBOHYD 889 889 O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT CARBOHYD 981 981 O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT CARBOHYD 1763 1763 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2206 2206 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2528 2528 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2679 2679 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 25 1188 Missing (in isoform Short).
SQ SEQUENCE 3063 AA; 333194 MW; 75FEA78FA8E48293 CRC64;
```

Alignment Scores:
Pred. No.: 3,16e-11 Length: 3063
Score: 261.50 Matches: 64
Percent Similarity: 52.5% Conservative: 29
Best Local Similarity: 36.2% Mismatches: 83
Query Match: 24.9% Indels: 1
DB: 1 Gaps: 1

US-10-699-035A-1 (1-537) x COCAI_HUMAN (1-3063)

```
QY 4 GACCTGATGTTCTCTGCTGAGACAGAGCTGCTCTCACTACGAGTTCTCCGGATT 63
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 440 AspIleValPheLeuValAspGlySerTyrSerIleGlyIleAlaSnPheValIysVal 459
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 64 CGGGAGTTTGTGGGGCAGCTGTGCTCACTGCGCCCGGGCCCTGCGTGC 123
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 460 ArgAlaPheLeuGluValLeuValIysSerPheGluIleSerProAsnArgValGlnIle 479
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 124 AGCTGTGTCACGTGGGCGACATGCGCATACCGAGTCCCTTGGCCGACGACACTCG 183
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 480 SerLeuValGlnTyrSerArgAspProIleThrGluPheThrLeuIysPheThrIys 499
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 184 GGTGAGCGTCCGCGAGATGCGGTGCTTCCGCGGCGATGGGTGACACCCACT 243
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 500 ValGluAspIleIleGluAlaIleAsnThrPheProTyrArgGlySerThrAsnThr 519
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 244 GGCCTGGCGCTGTATGCAAGAACGCTGTTTCTGAAAGCATCAGGTGCCGCGCA 303
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 520 GlyLysAlaMetThrTyrValArgGluIysIlePheValProSerIysGlySerArgSer 539
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 304 GGGGTGCCCAAGTGTGTTGGTGAACATGGCGGCTCCAGCGACCTGTGGCCCC 363
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 540 AsnValProLysValMetIleLeuIleThrAspGlyLysSerSerAspAlaPheArgAsp 559
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 364 CCCATGACAGAGCTCAAGACCTGGGCGTCAACCGCTTCATTGTGACACCGCGCAGGC 423
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 560 ProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValIysAspAla 579
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 424 AACTTCTGAGCTGTACGCGCTGCTAGCCCTGCGGAGACCTGCATTTGTG 483
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 580 ValArgSerGluLeuGluAlaIleAlaSerProAlaGluThrIleValPheThrVal 599
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 484 ---GACGTGATGATCTGCACATCATTTGCAAGAGCTGAGGGGCTTCATT 531
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 600 GluAspPheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIle 616
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:12:53 ; Search time 18.5996 Seconds
(without alignments)
2538.475 Million cell updates/sec

Title: US-10-699-035A-1

Perfect score: 1049
Sequence: 1 999gagctcgtatctctcgtc.....tga99ggtccattctgcgcg 537

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 433378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-O=/abs/ABSSWB pool/US10699035/runat_13022006_062440_25403/app_query.fasta_1
-DB=A.Geneseq -GPM=factan -SUFFIX=rag -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=Pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=Pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=US10699035 @CGN 1.1 476 @runat_13022006_062440_25403 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.Geneseq.21.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*
- 9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	902	86.0	180	6	AAE32500	AAE32500 Human von
2	902	86.0	215	4	AAAB7344	AAAB7344 Human gen
3	902	86.0	215	5	ABG65347	ABG65347 Human alb
4	902	86.0	215	8	ADL78614	ADL78614 Albunin f
5	902	86.0	226	4	AAAB7424	AAAB7424 Human gen
6	902	86.0	242	4	AAAB7418	AAAB7418 Human gen
7	902	86.0	285	9	AAAB8340	AAAB8340 Human mem
8	902	86.0	285	9	ADY63045	ADY63045 Human c10
9	902	86.0	418	6	AAE32502	AAE32502 Human Wtl

10	902	86.0	445	5	ABP69674	ABP69674 Human pol
11	902	86.0	445	8	ADH71106	ADH71106 Human pro
12	886	84.5	185	4	AAE03654	AAE03654 Human ext
13	877	83.6	299	3	AAAB42581	AAAB42581 Human ORF
14	872	72.6	180	6	AAE32503	AAE32503 Mouse von
15	762	72.6	415	6	AAE32501	AAE32501 Mouse Wtl
16	754	71.9	421	2	AAAB6326	AAAB6326 Kidney in
17	751	71.6	186	4	AAAB7419	AAAB7419 Human gen
18	384	36.6	77	3	AAAG02196	AAAG02196 Human sec
19	272.5	26.0	176	5	ABAB78816	ABAB78816 Von Wille
20	265	25.3	644	8	ADP66963	ADP66963 Mouse can
21	265	25.3	644	8	ADL12632	ADL12632 Murine can
22	261.5	24.9	3063	5	ABE90762	ABE90762 Human Tum
23	261.5	24.9	3063	6	ABU54469	ABU54469 Human tum
24	261.5	24.9	3063	6	ABR47415	ABR47415 Breast ca
25	261.5	24.9	3063	6	ABR47416	ABR47416 Breast ca
26	261.5	24.9	3063	6	ADJ75666	ADJ75666 Marker ge
27	261.5	24.9	3118	4	AAU27790	AAU27790 Human ful
28	259	24.7	293	3	AAAB39142	AAAB39142 Human sec
29	259	24.7	357	4	AAAB93189	AAAB93189 Human pol
30	259	24.7	357	8	ADL30530	ADL30530 Human pro
31	259	24.7	391	9	ADL12639	ADL12639 Human can
32	259	24.7	488	4	AAAB93707	AAAB93707 Human pol
33	259	24.7	488	8	ADL31608	ADL31608 Human pro
34	259	24.7	499	9	ADZ12635	ADZ12635 Human can
35	259	24.7	499	9	ADZ12643	ADZ12643 Human can
36	259	24.7	540	9	ADZ12637	ADZ12637 Human can
37	259	24.7	540	9	ADZ12641	ADZ12641 Human can
38	259	24.7	580	8	ADQ19949	ADQ19949 Human sof
39	259	24.7	581	8	ADP66966	ADP66966 Human can
40	259	24.7	581	9	ADZ12645	ADZ12645 Human can
41	259	24.7	581	9	ADZ12647	ADZ12647 Human can
42	259	24.7	620	9	ADZ12649	ADZ12649 Human can
43	254.5	24.3	795	5	AAU84267	AAU84267 Human end
44	254.5	24.3	795	4	AAAB27229	AAAB27229 Human EXM
45	254.5	24.3	1207	7	ADP69785	ADP69785 Human hea

ALIGNMENTS

RESULT 1	AAE32500	standard, protein, 180 AA.
ID	AAE32500	standard, protein, 180 AA.
XX	AAE32500;	
AC	AAE32500;	
XX	AAE32500;	
DT	24-MAR-2003	(first entry)
XX	24-MAR-2003	(first entry)
DE	Human von Willebrand Factor A (VA) domain.	
XX	Human von Willebrand Factor A (VA) domain.	
KW	Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;	
KW	extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;	
KW	gene therapy; human.	
XX	gene therapy; human.	
OS	Homo sapiens.	
XX	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	Misc-difference 1..2	
FT	Misc-difference 179..180	/note="Encoded by GGG"
FT	Misc-difference 179..180	/note="Encoded by CTC"
FT	Misc-difference 179..180	/note="Encoded by CTC"
XX	W0200288184-A1.	
PN	W0200288184-A1.	
XX	W0200288184-A1.	
PD	07-NOV-2002.	
XX	07-NOV-2002.	
XX	02-MAY-2002; 2002W0-AU000542.	
PF	02-MAY-2002; 2002W0-AU000542.	
XX	02-MAY-2002; 2002W0-AU000542.	
PR	02-MAY-2002; 2002W0-AU000542.	
XX	02-MAY-2002; 2002W0-AU000542.	
XX	(MURD-) MURDOCH CHILDRENS RES INST.	
PA	(MURD-) MURDOCH CHILDRENS RES INST.	
XX	(MURD-) MURDOCH CHILDRENS RES INST.	
PI	Bateman JF, Fitzgerald DJ;	


```
XX      Sequence 215 AA;
SQ
Alignment Scores:
Pred. No.:      4,686-69      Length:      215
Score:          902.00      Matches:      178
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    86.0%      Indels:      0
DB:              Gaps:      0
US-10-699-035A-1 (1-537) x AAB87344 (1-215)

QY      1 GGGGACCTGATGTTCTCTGTGACAGCTGACCCAGCGTCTCTACATACAGATTCTCCCG 60
DB      33 GtYAspLeuMeTPhelLeuLeuAspSerSerAlaSerValSerHisIryGluPheSerArg 52
QY      61 GTTCCGGAGTTGTGGGGCAGCTGTGTGCTCCACTGCCCCCTGGGCAACCGGGGCTTGGCGT 120
DB      53 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY      121 GGCAGCTGTGACGTGGGGAGTGGGCCATACAGGAGTCCCTTGGCCAGCACAGC 180
DB      73 AlaSerLeuValHisValGlnSerArgProIryThrGluPheProPheGlyGlnHisSer 92
QY      181 TCGGGTGAAGCTGCCCCAGGATGCGGTGCTTCTGCCCCAGCGCATGGGTGACACCCAC 240
DB      93 SerGlyGluAlaAlaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlyAspThrHis 112
QY      241 ACTGGCCTGGCGCTGTGTTATGCCAAGAAACAGCTGTTTGTGAAGCATCAGTGCCTGG 300
DB      113 ThrGlyLeuAlaLeuValIryAlaIryGlnLeuPheAlaGlnAlaSerGlyAlaArg 132
QY      301 CCAAGGGGTGCCCCAAAGTGTGTGTGGGTGACAGATGGGGGTCCAGCGACCTGAGGGC 360
DB      133 ProGlyValProIryValLeuValIryPvalThrAspGlyLysSerSerAspProValGly 152
QY      361 CCCCCCATGACGAGCTCAAGACCTGGGCGTCAACCGTGTTCATTGTACACCGGCGCA 420
DB      153 ProPheMetGlnGlnLeuLeuAspLeuGlyValThrValPheIleValSerThrGlyArg 172
QY      421 GGGCACTTCCCTGGAGCTGTGACGCGCTGCTCCAGCCCTGCGGAGAAACACTGCACTTT 480
DB      173 GlyAspPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 192
QY      481 GTGGAGCTGATGACCTGCACATCATTTGCCAAGAGCTGAGGGGCTCCATTCTC 534
DB      193 ValAspValAlaAspAspLeuHisIleIleValGlnGlnLeuArgGlySerIleLeu 210

RESULT 3
ABG65347
ID      ABG65347 standard; protein; 215 AA.
AC      ABG65347;
XX
DT      27-AUG-2002 (first entry)
XX
DE      Human albumin fusion protein #2022.
XX
KW      Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW      human serum albumin; HSA; cancer; reproductive disorder;
KW      digestive disorder; immune disorder; endocrine disorder;
KW      haematopoietic disorder; neural disorder; connective disorder;
KW      cytotoxic; antiinfectivity; antiinflammatory; anticancer;
KW      immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
KW      neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW      osteopathic; antiarthritic.
OS      Homo sapiens.
OS      Synthetic.
XX
XX      WO200177137-A1.
XX
```

```
PD      18-OCT-2001.
XX
XX      12-APR-2001; 2001WO-US011988.
XX
XX      12-APR-2000; 2000US-0229358P.
PR      25-APR-2000; 2000US-0199384P.
PR      21-DEC-2000; 2000US-0256931P.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Rosen CA, Haeseltine WA;
XX
XX      WPI; 2002-010886/01.
XX
XX      New fusion protein for treating disease e.g. diabetes comprises an
XX      albumin fused to a therapeutic protein.
XX
XX      Claim 1; Page 1935; 2102pp; English.
XX
XX      The present invention relates to albumin fusion proteins comprising a
XX      therapeutic protein X and human albumin (HA, also known as human serum
XX      albumin, HSA). The proteins are useful for treating a disease or disorder
XX      that may be modulated by therapeutic protein X. The albumin extends the
XX      shelf-life of protein X, and may increase its biological in vitro/in vivo
XX      activity. The protein is useful for treating and diagnosing disorders
XX      such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
XX      disease, ulcerative colitis), immune disorders (e.g. acquired
XX      immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
XX      haematopoietic disorders, neural disorders (e.g. Alzheimer's,
XX      Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
XX      schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).
XX      ABG63326-ABG65518 represent albumin fusion proteins of the invention
XX
SQ      Sequence 215 AA;
Alignment Scores:
Pred. No.:      4,686-69      Length:      215
Score:          902.00      Matches:      178
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    86.0%      Indels:      0
DB:              Gaps:      0
US-10-699-035A-1 (1-537) x ABG65347 (1-215)

QY      1 GGGGACCTGATGTTCTCTGTGACAGCTGACCCAGCGTCTCTACATACAGATTCTCCCG 60
DB      33 GtYAspLeuMeTPhelLeuLeuAspSerSerAlaSerValSerHisIryGluPheSerArg 52
QY      61 GTTCCGGAGTTGTGGGGCAGCTGTGTGCTCCACTGCCCCCTGGGCAACCGGGGCTTGGCGT 120
DB      53 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY      121 GGCAGCTGTGACGTGGGGAGTGGGCCATACAGGAGTCCCTTGGCCAGCACAGC 180
DB      73 AlaSerLeuValHisValGlnSerArgProIryThrGluPheProPheGlyGlnHisSer 92
QY      181 TCGGGTGAAGCTGCCCCAGGATGCGGTGCTTCTGCCCCAGCGCATGGGTGACACCCAC 240
DB      93 SerGlyGluAlaAlaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlyAspThrHis 112
QY      241 ACTGGCCTGGCGCTGTGTTATGCCAAGAAACAGCTGTTTGTGAAGCATCAGTGCCTGG 300
DB      113 ThrGlyLeuAlaLeuValIryAlaIryGlnLeuPheAlaGlnAlaSerGlyAlaArg 132
QY      301 CCAAGGGGTGCCCCAAAGTGTGTGTGGGTGACAGATGGGGGTCCAGCGACCTGTTGGC 360
DB      133 ProGlyValProIryValLeuValIryPvalThrAspGlyLysSerSerAspProValGly 152
QY      361 CCCCCCATGACGAGCTCAAGACCTGGGCGTCAACCGTGTTCATTGTACACCGGCGCA 420
DB      153 ProPheMetGlnGlnLeuLeuAspLeuGlyValThrValPheIleValSerThrGlyArg 172
```

QY	421	GGCACTTCTCTGGAGCTGTTCAGCCCGTGTGCTCTACGCCCTTCGACGAAGACACTGTGACTTT	480
DB	173	GIYAAPHPELENGCLUENUSERIAAIAAIAISERLAProAlaGluIuYshISleuHISpe	192
QY	481	GTGGAAGTGTGAAGACTGTGCACATCATTTGTCGAAGAGCTGAGGGGCTCCATTTTC	534
DB	193	ValAspValAlAspAspIeuHISleIleValGlnGlnIuIeuArgGlyserIleIeu	210
RESULT 4			
ADL78614			
ID	ADL78614	standard; protein; 215 AA.	
XX	AC	ADL78614;	
XX	DT	20-MAY-2004 (first entry)	
DE	XX	Albumin fusion protein related therapeutic protein X, SEQ ID NO 2096.	
KM	KM	albumin fusion protein; cytostatic; antianaemic; antiarthritic;	
KM	KM	antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;	
KM	KM	antipsoiatric; antibacterial; osteopathic; dermatological; antigout;	
KM	KM	immunomodulator; antiarrhythmic; cardiac; nootropic; antilipemic;	
KM	KM	nephroretropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;	
KM	KM	antidiabetic; anabolic; hypertensive; vulnery; gene therapy; cancer;	
KM	KM	reproductive system disorder; therapeutic protein.	
OS	XX	Unidentified.	
PN	XX	US2004010134-A1.	
PD	XX	15-JAN-2004.	
PF	XX	12-APR-2001; 2001US-00833245.	
PR	XX	12-APR-2000; 2000US-0229358P.	
PR	XX	25-APR-2000; 2000US-0199384P.	
PR	XX	21-DEC-2000; 2000US-0256931P.	
PA	XX	(ROSE/) ROSEN C A.	
PA	XX	(HASE/) HASELTINE W A.	
PI	XX	Rosen CA, Haseeltine WA;	
DR	XX	WPI; 2004-090519/09.	
PT	XX	New albumin fusion proteins, useful for diagnosing, treating, preventing	
PT	XX	or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,	
PT	XX	asthma, inflammatory bowel disease or Alzheimer's disease.	
PS	XX	Disclosure; SEQ ID NO 2096; 279pp; English.	
CC	CC	The invention relates to a novel albumin fusion protein. The invention	
CC	CC	further relates to: a composition comprising the albumin fusion protein	
CC	CC	and a pharmaceutical carrier; a kit comprising the composition of the	
CC	CC	albumin fusion protein formula; a method of treating a disease or	
CC	CC	disorder in a patient comprising the step of administering the albumin	
CC	CC	fusion protein; a method of treating a patient with a disease or disorder	
CC	CC	that is modulated by Therapeutic protein: X, or its fragment or variant;	
CC	CC	a method of extending the shelf life of Therapeutic protein: X, or its	
CC	CC	fragment or variant; a nucleic acid molecule comprising a polynucleotide	
CC	CC	sequence encoding the albumin fusion protein; a vector comprising the	
CC	CC	nucleic acid molecule of the albumin fusion protein; and a host cell	
CC	CC	comprising the nucleic acid molecule of the albumin fusion protein. The	
CC	CC	albumin fusion protein and its compositions have the following	
CC	CC	activities: cytostatic, antianaemic, antiarthritic, antiasthmatic, anti-	
CC	CC	HIV, immunosuppressive, antiinflammatory, antipsoiatric, antibacterial,	
CC	CC	osteopathic, dermatological, antigout, immunomodulator, antiarrhythmic,	
CC	CC	cardiac, nootropic, antilipemic, nephroretropic, uropathic,	
CC	CC	neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic,	
CC	CC	hypertensive, and vulnery. The albumin fusion protein nucleic acid may	
CC	CC	be used in gene therapy to treat disorders. The albumin fusion protein is	
CC	CC	useful for diagnosing, treating, preventing or ameliorating diseases or	
CC	CC	disorders comprising indication: Y. The diseases or disorders include:	

CC	cancer (e.g. leukaemia, colon, bone, breast, liver or lung cancer),
CC	immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute
CC	lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS,
CC	autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
CC	disease), reproductive system disorders (e.g. prostaticitis, inguinal
CC	hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
CC	leydig tumours), musculoskeletal diseases (e.g. giant cell tumours,
CC	Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy
CC	or cachexia), cardiovascular disease (e.g. rhadomyomas, heart disease,
CC	arrhythmia), cardiac arrest, heat valve disease, hypernatraemia or
CC	hyponatremia), mixed foetal diseases (e.g. foetal alcohol syndrome,
CC	Down's syndrome, Patau syndrome, Turner's syndrome, Apert syndrome or Tay-
CC	-Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
CC	tract infections or renal disorders), neural or sensory disease (e.g.
CC	Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
CC	cerebellar ataxia, attention deficit disorder, autism or obsessive
CC	compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or
CC	occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
CC	disease or glomerulonephritis), digestive diseases (e.g. portal
CC	hypertension, irritable bowel disease, gastric atrophy or pancreatitis)
CC	or connective tissue or epithelial diseases (e.g. Crohn's disease,
CC	scleroderma, wound healing or epidermolysis bullosa). This sequence
CC	represents a therapeutic protein X relating to the albumin fusion protein
CC	of the invention. The sequence listing data for this specification was
CC	downloaded from the USPTO website.
XX	
SQ	Sequence 215 AA;
Alignment Scores:	
Pred. No.:	4.68e-69 Length: 215
Score:	902.00 Matches: 178
Percent Similarity:	100.0% Conservative: 0
Best Local Similarity:	100.0% Mismatches: 0
Query Match:	86.0% Indels: 0
DB:	8 Gaps: 0
US-10-699-035A-1 (1-537) x ADL78614 (1-215)	
OY	1 GGGAAGCTGATGGTTCCTGTGGACAGCTCAGCCAGCGTTCTCACTAGAGATTCTCCCG 60
Db	33 GlyaaPLeuMeIrrPhleuLeuabserSerlaIaservaISeriHirfoIuhesaraYg 52
OY	61 GTTGCGGAGTTTGTGGGCGAGCTGGTGGCTCCACTGCCCTTGCGCACCGGGGCGCTGGCT 120
Db	53 ValaRgiuPheValIGlYInleuValaIAProIeuPProIeuGlYThrgIyalaueaYg 72
OY	121 GCCATCTGGTGCACAGTGGCGCATGCAGCATCACCGAGTTCCCTTGGCGGCACAGC 180
Db	73 AIservleuValHisIvalGISeIxrPrOIYrThngIuPhePrObelYglNhIsSer 92
OY	181 TCGGGTGAAGCTGCCAGAGATCGGTGGTGGCTTGTGCCCAAGCGATGGGTGATACACCAC 240
Db	93 SerGIyGIuaIalalagIlnsPlalValaRgalaserlaIagIlnRgmecIyasnTrnHIs 112
OY	241 ACTGGCTGGCGGCTGCTATGCCAAGAACAAGCTGTTTGTGMAAGATCAGGTGCCGG 300
Db	113 ThrgIyaleuAlaleuValYrZalalySGluGInleuPhealIGlUalaseRgIyalaYg 132
OY	301 CCAGGGGTGCCAAAGTGTGGTGGTGGGAGAACATGGCGGCTCCAGAGCAACCTTGGGGC 360
Db	133 ProGIyValProlYsValIleuValITryValITrnaspGIYGIYSerSerIasPProValIGly 152
OY	361 CCCCCATGACAGAGACTCAAAGACCTGGGGCGCACCGTTCATTATGTTCAGACCGGCGGA 420
Db	153 ProPProMetGIingIInleuYaSPleuGIYalITrnValPheIIleValSerThrgIyalaYg 172
OY	421 GGCAATCTTCCTGGAGCTGTCAACCGCTGTGCTTAGCCCTCCAGAAAGCAACCTGTGACTTT 480
Db	173 GIyaanPheIreGIuIeuserIrlaIalaserlaIaProlaIGlUylshIsIeunHIsPhe 192
OY	481 GTGAAGCTGGATGACCTGCATCATCTTGTCCAAGACTGAGGGGCTCCATTCTC 534
Db	193 ValaPValaAsPaSPleuHIsIleIleValaGIingIuIeuaRGIYSerIleuN 210

RESULT 5
 AAB87424
 ID AAB87424 standard; protein; 226 AA.
 XX
 AC AAB87424;
 XX
 DT 22-MAY-2001 (first entry)
 DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnery; cell culture;
 KW chemotaxis; food additive; binding partner identification.
 XX
 OS Homo sapiens.
 XX
 MN WO200118022-A1.
 XX
 PD 15-MAR-2001.
 XX
 PF 31-AUG-2000; 2000WO-US024008.
 XX
 PR 03-SEP-1999; 99US-0152315P.
 XX
 PR 03-SEP-1999; 99US-0152317P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsu GA, Rosen CA,
 PI Soppet DR, Young PE, Edner R, Duan DR, Olsen HS, Lafleur DW;
 PI Moore PA, Shi Y, Wei Y, Florence KA;
 XX
 DR WPI; 2001-203081/20.
 XX
 PT Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers.
 XX
 PT Disclosure; Page 18; 607pp; English.
 XX
 PS AAFP1856-AAFP1929 represent cDNAs corresponding to 52 human secreted
 CC protein genes, and AAB87412-AA87413 represent the proteins they encode.
 CC AAB87414-AA87415 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 52 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
 CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and

CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein fragment referred to in the disclosure of the invention
 XX
 SQ Sequence 226 AA;
 XX
 Alignment Scores:
 Pred. No.: 4,73e-69 Length: 226
 Score: 902.00 Matches: 178
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 86.0% Indels: 0
 DB: 4 Gaps: 0
 XX
 US-10-699-035A-1 (1-537) x AAB87424 (1-226)
 QY 1 GGGAGCCGATGTTCTCTGAGCGCTCAGCCAGCGCTCTCAATGAGTTCCCGG 60
 DB 46 GYAAPLeuMetPheLeuLeuAspSerSerAlaSerValSerHisArgLupheserArg 65
 QY 61 GTTCGGAGTTTGTGGGACAGCTGGGCTCCAGTCCGCGGACCGGGGCGTGGCGT 120
 DB 66 VALArgLupheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 85
 QY 121 GCCAGCTGTGACAGTGGGAGTGGGCGCATACACCGAGTTCCCTGGCCAGCACAGC 180
 DB 86 AlaSerLeuValHisValGlySerArgProThrGlnPheProPheGlyGlnHisSer 105
 QY 181 TCGGTTGAGGCTGCCAGAGATGCGGTGCTTCTGCGCCAGCGATGGGTACACCCAC 240
 DB 106 SerGlyGlnAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 125
 QY 241 ACTGCGCTGGCGCTGTCTATGCCAAGAACAGCTGTTGCTGAAGCATCAGAGTCCCGG 300
 DB 126 ThrGlyLeuAlaLeuValTyrAlaGlySerGlnLeuPheAlaGlnHisSerGlyAlaArg 145
 QY 301 CCAGGCGTGGCCAAAGTCTGTGTGGTGCAGATGGCGGCTCCAGCGACCTGTGGGC 360
 DB 146 ProGlyValProLysValLeuValThrValThrAspGlyGlySerSerAspProValGly 165
 QY 361 CCCCCCATGGAGAGCTCAAGACCTGGGCGCTCAACCGTTCATGTCAGACCGCGCA 420
 DB 166 ProPheMetGlnGlnLeuLysPheuGlyValThrValPheIleValSerThrGlyArg 185
 QY 421 GGCAACTCTCGAGAGCTGTCAAGCCGCTCCAGCGCCGCGAGAACACCTGCATTT 480
 DB 186 GYAenPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 205
 QY 481 GTGACGTGATGATGACCTGCACATCATTTGCCAAGAGTGGGCGCTCCATTTCTC 534
 DB 206 VALAspValAspAspLeuHisIleIleValGlnGlnLeuArgGlySerIleLeu 223
 XX
 RESULT 6
 AAB87418
 ID AAB87418 standard; protein; 242 AA.
 XX
 AC AAB87418;
 XX
 DT 22-MAY-2001 (first entry)
 DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnery; cell culture;
 KW chemotaxis; food additive; binding partner identification.
 XX

OS Homo sapiens.
 XX WO200118022-A1.
 XX 15-MAR-2001.
 PD 31-AUG-2000; 2000WO-US024008.
 XX 03-SEP-1999; 99US-0152315P.
 PR 03-SEP-1999; 99US-0152317P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
 PI Soppe DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
 PI Moore PA, Shi Y, Wei Y, Florence KA;
 XX WPI; 2001-203081/20.
 DR Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers.
 PS Disclosure: Page 18, 607bp; English.
 XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
 CC protein genes, and AAB87342-AA87413 represent the proteins they encode.
 CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 52 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
 CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunoassorbent assay (ELISA). The present sequence represents a human
 CC secreted protein fragment referred to in the disclosure of the invention
 XX
 XX SQ Sequence 242 AA;
 Alignment Scores:
 Pred. No.: 4 79e-69 Length: 242
 Score: 902.00 Matches: 178
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 86.0% Indels: 0
 DB: 4 Gaps: 0
 US-10-699-035A-1 (1-537) x AAB87418 (1-242)
 QY 1 GGGGACCTGATGTTCTGCTGGACAGCTGACGAGCTCTCTACATGAGTTCCTCCGG 60
 Db 60 GtlyAspIeueMetPheIeueIueNspSersSerAlaSerValSerHiserylIuheserArg 79
 QY 61 GTTCGGAGATTGTGTGGGGCAAGCTGTGCTCCACTGCCCCCTGGGCAACCGGGCCCTGCGT 120

Db 80 ValArgIuPheValGIgIuIeueValAlaProIeueProIeueGIyThrGIyAlaIeueArg 99
 QY 121 GCCAGCTGTGTGCAAGTGGGCAAGTCGGCCATACACCGATTCCCTTGGCCAGACAGC 180
 Db 100 AlaSerIeueValHisValGIySerArgProIyThrGIuPheProPheGIyIuhIser 119
 QY 181 TCGGGTGAAGGCTGCCCAAGATGCGGTGCTCTTCTGCGCCAGCGCATGGGTGACACCCAC 240
 Db 120 SerGIyIuAlaAlaGIuIeueValAlaArgAlaSerAlaGIuIeueMetGIyAspThrHis 139
 QY 241 ACTGGCCCTGGCGCTGCTATGCTCCAGGAAACAGCGTGTTCGTAAGCATTCAGTGCCCGG 300
 Db 140 ThrGIyIeueAlaIeueValTyrAlaIySerGIuIeuePheAlaGIuIeueSerGIyAlaArg 159
 QY 301 CCAGGGGTGCCCAAGT 360
 Db 160 ProGIyValProIyValProIyValProIyValProIyValProIyValProIyValProIyVal 179
 QY 361 CCCCCATGCAAGAGCTCAAGACCTGGGCGTCAACCGTTCATTGTCATGTCAGCACCGCCGA 420
 Db 180 ProPheMetGIuIeueValAspIeueGIyValThrValPheIleValSerThrGIyArg 199
 QY 421 GGGCACTTCCGTGAGCTGTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 Db 200 GtlyAspIeueIuIeueValSerAlaAlaIaSerAlaProAlaGIuIyHisIeueIspHe 219
 QY 481 GTGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 534
 Db 220 ValAspValAspAspIeueHisIleIleValGIuIeueArgGIySerIleIeue 237
 RESULT 7
 AAB88340
 ID AAB88340 standard; protein, 285 AA.
 AC AAB88340;
 DT 23-MAY-2001 (first entry)
 XX Human membrane or secretory protein clone PSEC0053.
 DE Human membrane or secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes.
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 OS Homo sapiens.
 XX EPI067182-A2.
 PN EPI067182-A2.
 PD 10-JAN-2001.
 XX 07-JUL-2000; 2000EP-00114090.
 PF 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 FR 02-MAY-2000; 2000JP-00183766.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isegai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 PI WPI; 2001-093889/11.
 DR N-PSDB; AAF93767.
 XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development.
 PT Claim 1; SEQ ID NO 48; 609BP + Sequence listing; English.
 PS This invention relates to nucleic acid sequences AAF93744 - AAF93916
 XX which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF92232 - AAF62235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of cDNA
 CC antibodies directed against the proteins, and cDNA sequences, which can

be used in vaccines. The polynucleotide sequences can be used in gene therapy. The polynucleotide sequences and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbent assay (ELISA). Examples of diseases which may be treated include rheumatoid arthritis and diabetes

XX Sequence 285 AA;

Alignment Scores:
Pred. No.: 4,95e-69 Length: 285
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 4 Gaps: 0

US-10-699-035a-1 (1-537) x AAB88340 (1-285)

QY 1 GGGAGCTGATGTTCTGCTGAGACGCTCAGCCAGCGTCTTCTACAGATTCTCCCGG 60
DB 33 GYAAPLeuMePheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCCGAGATTGTTGGGGAGCTGGGCTGCTCACTCCCTGGGACCGGGCCCTGGCT 120
DB 53 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY 121 GCCAGTCTGGTCACTGGGAGTCCGATACACCGAGTTCCTCCCTGGGACGACAC 180
DB 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGGATGCGGTGCGTCTTCTGCCAGCGCATGGGTGAACCCAC 240
DB 93 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 112
QY 241 ACTGCGCTGGCGCTGCTATGCTCAAGAAACAGCTGTTGCTGAGCATCAGTCCCGG 300
DB 113 ThrGlyLeuAlaLeuValTyrAlaTyrGlnGlnLeuPheAlaGlnAlaSerGlyAlaArg 132
QY 301 CCAGGGGTGCCCAAGTGTGGTGTGGTGAAGATGGGCGCTCCAGCGACCTGTGGGG 360
DB 133 ProGlyValProLysValLeuValTyrValThrAspGlyGlySerSerAspProValGly 152
QY 361 CCCCCCATGAGAGCTCAAGACCTGGGCGTCAACCGGTTCATTGTCAACCCGCCGA 420
DB 153 ProPheMetGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 172
QY 421 GCGCACTTCTGAGAGTGTCAAGCGCTGCTCAGCCCTCCGAGAAAGACCTGCACTT 480
DB 173 GlyAspPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 192
QY 481 GTGAGCTGATGATGCTGACATCATTTGCAAGAGCTGAGGGGCTCCATTCTC 534
DB 193 ValAspValAspAspLeuHisIleIleValGlnGlnLeuArgGlySerIleLeu 210

RESULT 8

ADY63045

ID ADY63045 standard; protein; 285 AA.

XX ADY63045;

DT 02-JUN-2005 (first entry)

XX Human clone PSEC0053 protein, SEQ ID 48.
DE Gene therapy.
KW Homo sapiens.
OS EP1514933-A1.
PN 16-MAR-2005.
PD 07-JUL-2000; 2004EP-00027228.
PF 08-JUL-1999; 99JP-00194179.
PR 11-JAN-2000; 2000JP-00118775.
PR 02-MAY-2000; 2000JP-00183766.
PR 07-JUL-2000; 2000EP-00114090.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Ota T, Isegai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI MPI: 2005-203865/22.
DR N-PSDB; ADY63044.

PT Novel isolated polynucleotide encoding human secretory proteins or
PT membrane proteins, useful for examination and diagnosis of abnormality of
PT human secretory proteins.
PS Disclosure; SEQ ID NO 48; 1240pp; English.

XX The present invention relates to novel human secretory proteins or
XX membrane proteins, and their coding sequences. The present sequence is
XX one such protein sequence. The coding sequences of the invention are
XX useful for examination and diagnosis of abnormality of the human
XX secretory proteins and in gene therapy methods. The coding sequences and
XX proteins are useful as candidates for medicines or as target molecules
XX for developing medicines. Antibodies against the proteins of the
XX invention are useful for treating diseases that are associated with the
XX proteins. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained from sequence information
XX supplied by the European Patent Office.

SQ Sequence 285 AA;

Alignment Scores:
Pred. No.: 4,95e-69 Length: 285
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 9 Gaps: 0

US-10-699-035a-1 (1-537) x ADY63045 (1-285)

QY 1 GGGAGCTGATGTTCTGCTGAGACGCTCAGCCAGCGTCTTCTACAGATTCTCCCGG 60
DB 33 GYAAPLeuMePheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCCGAGATTGTTGGGGAGCTGGGCTGCTCACTCCCTGGGACCGGGCCCTGGCT 120
DB 53 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY 121 GCCAGTCTGGTCACTGGGAGTCCGATACACCGAGTTCCTCCCTGGGACGACAC 180
DB 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGGATGCGGTGCGTCTTCTGCCAGCGCATGGGTGAACCCAC 240
DB 93 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 112
QY 241 ACTGCGCTGGCGCTGCTATGCTCAAGAAACAGCTGTTGCTGAGCATCAGTCCCGG 300

Db 113 ThrGlyLeuAlaLeuValTyrAlaIySgluInLeuPheAlaGluAlaSerGlyAlaArg 132
Qy 301 CCAAGGGGTGCCCAAGTCTGTGTGGTGAACAATGGCGGCTCCAGGACCTGTGGGC 360
Db 133 ProGlyValProlYsValLeuValTrrValThraSpGlyGlySerSerAspProValGly 152
Qy 361 CCCCCATGACAGAGACTCAAGAGACTGGCGCTGACCCGTTTCATTGTACACACCGCGCA 420
Db 153 ProPmetGIngluLeuIlyAspLeuGlyValThrValPheIleValSerThrGlyArg 172
Qy 421 GGCAACTTCCTGGAGCTGTACAGCCGCTGACGCCCGCCGAGAGACCTGCACATT 480
Db 173 GlyAsnPheLeuGluLeuSerAlaAlaIaSerAlaProAlaGluIlyHisLeuHisPhe 192
Qy 481 GTGACGTGATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
Db 193 ValAspValAspAspLeuHisIleIleValGIngluLeuArgGlySerIleLeu 210
RESULT 9
AAE32502
ID AAE32502 standard; protein; 418 AA.
AC AAE32502;
DT 24-MAR-2003 (first entry)
XX Human Willebrand Factor A domain related-protein (WARP).
DE Human Willebrand Factor A domain related-protein (WARP).
KW Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
KW extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
KW gene therapy; human.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= Signal_peptide
FT Protein 19..418
FT /note= "Human mature WARP protein"
FT Modified-site 148
FT /note= "O-glycosylation site"
FT Misc-difference 210..211
FT /note= "Encoded by CTCGGC"
FT Modified-site 264
FT /note= "N-glycosylation site"
FT Modified-site 359
FT /note= "N-glycosylation site"
FT Modified-site 361
FT /note= "O-glycosylation site"
FT Disulfide-bond 369..393
XX
XX MO200288184-A1.
XX
XX PD 07-NOV-2002.
XX
XX PF 02-MAY-2002; 2002WO-AU000542.
XX
XX PR 02-MAY-2001; 2001AU-00004701.
XX
XX PA (MURD-) MURDOCH CHILDRENS RES INST.
XX
XX PI Bateman JF, Fitzgerald DJ;
XX
XX DR WPI; 2003-111873/10.
XX
XX DR N-PSDB; AAD50399.
XX
XX PT New isolated Willebrand Factor A-Related Protein polypeptide useful for
XX the manufacture of a medicament in the treatment of a disease condition
XX of the extracellular matrix, in particular arthritis.
XX
XX PS Claim 11; Page 76-78; 103pp; English.
XX
XX CC The invention relates to Willebrand Factor A domain related-protein

CC (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
CC superfamily of extracellular matrix (ECM) proteins. WARP is used as a
CC molecular marker, used for detecting a loss of ECM integrity in an animal
CC subject, monitoring repair, regeneration or other disease processes in an
CC animal subject and detecting a disease condition or a propensity for the
CC development of a disease condition in an animal subject. The invention is
CC useful for the manufacture of a medicament in the treatment of a disease
CC condition of the ECM. The disease condition involves the cartilage, and
CC is preferably arthritis. The invention is also used in gene therapy. The
CC present sequence is human WARP protein
XX
XX SQ Sequence 418 AA:
XX
XX Alignment Scores:
XX Pred. No.: 5,35e-69 Length: 418
XX Score: 902.00 Matches: 178
XX Percent Similarity: 100.0% Conservative: 0
XX Best Local Similarity: 100.0% Mismatches: 0
XX Query Match: 86.0% Indels: 0
XX DB: 6 Gaps: 0
XX
XX US-10-699-035A-1 (1-537) x AAE32502 (1-418)
Qy 1 GGGGACCTGATGTTCTGCTGGAACAGCTCAGCCGCTCTCACTACGAGTTCTCCCGG 60
Db 33 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
Qy 61 GTTCGGAGATTGTGGGAGACCTGATGCTCAGTCCCTCGGACCGGACCCCTGGCT 120
Db 53 ValArgGluPheValGlyGluLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
Qy 121 GCCAGTGTGTGACAGTGGGAGTGGCCATACACCGAGTTCCCTTGGCCAGACAGC 180
Db 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
Qy 181 TCGGATGAGCTGCCCGAGATCGGATGCTCTTCTCCCGGACCGAGTGGAGACCCAC 240
Db 93 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 112
Qy 241 ACTGGCTGGCGCGCTGTATGCCAAGAACAGCTGTTTGTGAAGCATCAGTGCCCGG 300
Db 113 ThrGlyLeuAlaLeuValTyrAlaIySgluInLeuPheAlaGluAlaSerGlyAlaArg 132
Qy 301 CCAAGGGGTGCCCAAGTCTGTGTGGTGAACAATGGCGGCTCCAGGACCTGTGGGC 360
Db 133 ProGlyValProlYsValLeuValTrrValThraSpGlyGlySerSerAspProValGly 152
Qy 361 CCCCCATGACAGAGACTCAAGAGACTGGCGCTGACCCGTTTCATTGTACACACCGCGCA 420
Db 153 ProPmetGIngluLeuIlyAspLeuGlyValThrValPheIleValSerThrGlyArg 172
Qy 421 GGCAACTTCCTGGAGCTGTACAGCCGCTGACGCCCGCCGAGAGACCTGCACATT 480
Db 173 GlyAsnPheLeuGluLeuSerAlaAlaIaSerAlaProAlaGluIlyHisLeuHisPhe 192
Qy 481 GTGACGTGATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
Db 193 ValAspValAspAspLeuHisIleIleValGIngluLeuArgGlySerIleLeu 210
RESULT 10
ABP69674
ID ABP69674 standard; protein; 445 AA.
XX
XX AC ABP69674;
XX
XX DT 20-JAN-2003 (first entry)
XX
XX DB Human polypeptide SEQ ID NO 1721.
XX
XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
XX cell-proliferative disorder; neurodegenerative disease; bacterial;
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;

XX	arthritis; cytosolic; immunomodulator; nocrotic; neuroprotective;		
KM	antiparkinsonian; antidiabetic; immunosuppressive; dermatological;		
KM	haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;		
KM	antiarthritic.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200270539-A2.		
XX			
PD	12-SEP-2002.		
XX			
PF	05-MAR-2002; 2002WO-US005095.		
XX			
PR	05-MAR-2001; 2001US-00799451.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;		
PI	Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;		
PI	Wehrman T, Wang J, Wang D, Drmanac RT;		
XX			
DR	WPI; 2002-759812/82.		
DR	N-PSDB; AB211891.		
XX			
PT	New polynucleotides comprising sequences assembled from expressed		
PT	sequence tags (ESTs), useful for treating cell-proliferative,		
PT	neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet		
PT	or coagulation disorders.		
XX			
PS	Claim 9; SEQ ID NO 1721; 1012pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated polynucleotide (I) comprising a		
CC	nucleotide sequence selected from any of 948 sequences (AB211119-		
CC	AB212066) or their mature protein coding portion, active domain coding		
CC	protein or complementary sequences. The polynucleotides are useful for		
CC	identifying expressed genes or for physical mapping of human genome. The		
CC	encoded polypeptides (AB268902-AB269849) are useful as molecular weight		
CC	markers, as a food supplement, for generating antibodies, in medical		
CC	imaging, screening and diagnostic assays and for treating cell-		
CC	proliferative disorders (cancer), neurodegenerative diseases (Parkinson's		
CC	or Alzheimer's disease), autoimmune diseases (multiple sclerosis,		
CC	diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,		
CC	platelet or coagulation disorders, wound, burne, incision, ulcers, liver		
CC	or lung fibrosis, infections (bacterial, viral, fungal, parasitic),		
CC	arthritis, etc. Note: The sequence data for this patent did not form part		
CC	of the printed specification, but was obtained in electronic format		
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX			
SO	Sequence 445 AA;		
SO			
Alignment Scores:			
Pred. No.:	5, 41e-69	Length:	445
Score:	902.00	Matches:	178
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	86.0%	Indels:	0
DB:	5	Gaps:	0
US-10-699-035A-1 (1-537) x ABP69674 (1-445)			
QY	1 GGGGACCTGATGTTCTCTGTGACAGCTCAGCCAGCGTCTCTACACTAGAGTTCTCCCG	60	
Db	33 G YAspRnuwterPheutleuAAspSerSerrAlaSerValSerHisrYgIuPhSerAr	52	
QY	61 GTTCGGAGCTTTGTGGGGAGCAGTGGGCTTCACATGCCCCCTGGGACCGGGGCGCTCGT	120	
Db	53 ValArgGluuHeValGlyGluuEuValAlaProIeuProLeuGlyTThGlyAlaLeuArg	72	
QY	121 GCACGTCGTGTGACGTCGGGAGTCGGGACATACACCGAGTTCCCTTCGGCCAGACAGC	180	
Db	73 AlaserIeuValHisValGlySerArArgProCYrThGluPheroPhegIyGlnHisSer	92	
QY	181 TCGGGTGAGGCTCCAGAGATGGCGTGGCTTCTGCCACAGCGATGGTGACACCCAC	240	

Dd		93	SeGcIyGIuAlAIAIGlnhpRAlaValArgrAlaSerAlaGInArgrMetGlyAerPhnHis	112
QY		241	ACTGCCTGGCGCTGCTGTATGCCAAGACAGCTGTTTGCTGAAGCATAGSTGCCCG	300
Dd		113	ThrGIyLeuAlaLeuValITyAlAlaySgInuInleuPheAlaGluAlaSerGIyAlaArg	132
QY		301	CCAGGGGTGCCCAAAAGTCGTGTGGTGGTGAACAATGGCGGCTCCAGAACCCGTGGGC	366
Dd		133	ProGIyValProLyValLeuValITryAlThrApGrIgylSerSerAerProValGIy	152
QY		361	CCCCCATTACAGACCTCAAGACCTGGGCGCTCACCCGTTCATGTACAGACCAGCCGA	420
Dd		153	ProPrometGIuInleuLysAspleucIyAlIThrValPheIleValSerThrGIyArg	172
QY		421	GCGCACTTCCTGAGACTGTCAAGCGGCTGCTCAAGCCCCTGCCGAAGAACACTGCACTTT	480
Dd		173	GLyAsnPheLeuGlutleuSerAlaAlaIaSerAlaProAlaGluLysHisLeuHisPhe	192
QY		481	GTGAGACGTGAGATGACCTGCACATCTTGTCCAAAGCTGAGAGGGCTCCATTCTC	534
Dd		193	ValAspValaAspAspleuHisIleIleValGIuInleuAArgIySerIleLeu	210
RESULT 11				
	ADH71106			
ID	ADH71106	standard; protein; 445 AA.		
XX				
AC	ADH71106;			
DT	25-MAR-2004	(first entry)		
XX				
DE		Human protein of the invention NOVA SEQ ID NO:2.		
KX		human; cytosolic; immunomodulator; neuroprotective; nootropic;		
KM		anorectic; antidiabetic; antimicrobial; antiinflammatory; gene therapy;		
KW		vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;		
KM		obesity; diabetes; infectious disease; metabolic syndrome X;		
KM		dyslipidaemia.		
XX				
OS	Homo sapiens.			
PN	WO2003102155-A2.			
PD	11-DEC-2003.			
XX				
Pf	03-JUN-2003;	2003WO-US017430.		
XX				
PR	03-JUN-2002;	2002US-0385120P.		
PR	04-JUN-2002;	2002US-0385784P.		
PR	05-JUN-2002;	2002US-0386041P.		
PR	05-JUN-2002;	2002US-0386047P.		
PR	06-JUN-2002;	2002US-0386376P.		
PR	06-JUN-2002;	2002US-0386453P.		
PR	06-JUN-2002;	2002US-0386644P.		
PR	06-JUN-2002;	2002US-0387016P.		
PR	07-JUN-2002;	2002US-0386796P.		
PR	07-JUN-2002;	2002US-0386931P.		
PR	07-JUN-2002;	2002US-0386942P.		
PR	07-JUN-2002;	2002US-0386971P.		
PR	07-JUN-2002;	2002US-0387162P.		
PR	08-JUN-2002;	2002US-0296960P.		
PR	10-JUN-2002;	2002US-0387400P.		
PR	10-JUN-2002;	2002US-0387535P.		
PR	11-JUN-2002;	2002US-0387610P.		
PR	11-JUN-2002;	2002US-0387625P.		
PR	11-JUN-2002;	2002US-0387634P.		
PR	11-JUN-2002;	2002US-0387668P.		
PR	11-JUN-2002;	2002US-0387696P.		
PR	11-JUN-2002;	2002US-0387702P.		
PR	11-JUN-2002;	2002US-0387836P.		
PR	11-JUN-2002;	2002US-0387859P.		
PR	12-JUN-2002;	2002US-0387933P.		

Human; extracellular matrix and cell adhesion molecule; XMAP;
 gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
 sickle cell anaemia; thalassemia; autoimmune disorder; adenocarcinoma;
 inflammatory disorder; acquired immune deficiency syndrome; AIDS;
 Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
 Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
 glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
 osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
 infection; cell proliferative disorder; actinic keratosis; myeloma;
 arteriosclerosis; neotropic; anticonvulsant; antithyroid; nephrotoxic;
 neuroprotective; dermatological.

Key Location/Qualifiers
 Region 1..171
 /note="Collagen glycoprotein precursor"
 Region 1..170
 /note="Von Willebrand factor domain score"
 Region 2..15
 /note="Collagen glycoprotein precursor"
 Region 37..51
 /note="Collagen glycoprotein precursor"
 Region 103..111
 /note="Collagen glycoprotein precursor"

MO200142285-A2.
 14-JUN-2001.
 05-DEC-2000; 2000WO-US032990.
 10-DEC-1999; 99US-0172852P.
 16-DEC-1999; 99US-0172354P.
 (INCY-) INCYTE GENOMICS INC.
 Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C,
 Baughn MR, Lu DM, Shah P, Au-Young J;
 WPI; 2001-381632/40.
 N-PSDB; AAD08062.

New human extracellular matrix and cell adhesion molecules and
 polynucleotide sequences encoding them, useful for diagnosis, prevention,
 treatment of genetic, autoimmune and cell proliferative disorders.

Claim 1; Page 111, 135pp; English.

The present sequence is a human extracellular matrix and cell adhesion
 molecule (XMAP). The XMAP is used for screening a compound for
 effectiveness as an agonist or antagonist of XMAP. The identified agonist
 or antagonist are used for treating a disease or condition associated
 with decreased or increased expression of functional XMAP. The
 polynucleotides encoding XMAP are useful in somatic or germline gene
 therapy to correct a genetic deficiency, to express a conditional gene
 lethal gene product and to express a protein which affords protection
 against intracellular parasites and also for diagnosis of disorders
 associated with expression of XMAP. They are also used for generating
 hybridisation probes useful in mapping the naturally occurring genomic
 sequences and to create knock in humanised animals (pigs) or transgenic
 animals (mice or rats) to model human diseases. Oligonucleotide or longer
 fragments derived from the polynucleotide sequences may be used as
 elements on a microarray. Antibodies which specifically bind XMAP may be
 used for the diagnosis of disorders associated with the expression of
 XMAP, or in assays to monitor patients being treated with XMAP. Diseases
 diagnosed, prevented or treated include genetic disorders such as
 adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 disease, myotonic dystrophy, sickle cell anaemia, thalassemia,
 autoimmune/inflammatory disorders such as acquired immune deficiency
 syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
 atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,

CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 bacterial, fungal, parasitic, protozoal and helminthic infections and
 cell proliferative disorders such as actinic keratosis, arteriosclerosis
 CC and cancer including breast, bladder, bone marrow, brain and uterus
 CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma
 XX

Sequence 185 AA;
 SQ

Alignment Scores:
 Pred. No.: 1,096-67 Length: 185
 Score: 886.00 Matches: 175
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 84.5% Indels: 0
 DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x AAE03654 (1-185)

QY 10 ATGTTCTGCTGAGACGCTCAGCCAGCGCTCTCTCACTACGAGTTCTCCGGGTTGGAG 69
 DB 1 MetPheLeuLeuAspSerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlu 20

QY 70 TTTGTGGGCGAGCTGTGGCTTCACTGCCCTGGGACACCGGGCCCTGCTGCCAGTCTG 129
 DB 21 PheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeu 40

QY 130 GTGCGCGGGGAGCGGCGGCGCATACACGAGTTCCCTCGGCGACGACGCTGGGCTGAG 189
 DB 41 ValHisValGlySerArgProTyrThrGlnPheProPheGlyGlnHisSerSerArgGlu 60

QY 190 GCTGCCAGATGCGGCGGCTGCTTCTGCGCAGCGCATGGGTGACACCCACACTGGCCTG 249
 DB 61 AlaAlaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeu 80

QY 250 GCGCTGCTCTATGCCAAGAACAGCTGTTTGTGAAGCATACAGTGGCCCGGACAGGGGTG 309
 DB 81 AlaLeuValTyrAlaTyrGlnGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyVal 100

QY 310 CCCAAGTGTGCTGTGGTGTGACAGATGGCGGCTCCAGCGACCTGGGGCCCCCCTATG 369
 DB 101 ProLysValLeuValTyrPheValThrAspLysValSerSerAspProValGlyProTromet 120

QY 370 CAGGAGCTCAAGGACCTGGGCGTCACTGTTTCATTTGTCAGGACCGGCGAGGCAACTTC 429
 DB 121 GlnGlnLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArgGlyLysPhe 140

QY 430 CTGAGCTGTACGCCGCTGCTCAGCCCTGCGGAGAACACCTGACCTTTGTGACGTG 489
 DB 141 LeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspVal 160

QY 490 GATGACCTGCACATCATTTGTCCAAAGAGCTGAGGGCTCCATTCTC 534
 DB 161 AspAspLeuHisIleIleValGlnIleuLysGlySerIleLeu 175

RESULT 13
 AAB42581
 ID AAB42581 standard; protein; 299 AA.
 AC AAB42581;
 XX
 XX 08-FEB-2001 (first entry)
 DE Human ORF2345 polypeptide sequence SEQ ID NO:4690.

Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
 vulnerability; antiportal; antiparkinsonian; neurotoxic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antiadrenergic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive.
 OS Homo sapiens.
 PN WO20058473-A2.
 XX 05-OCT-2000.
 XX 31-MAR-2000; 2000WO-US008621.
 PF 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX (CURA-) CURAGEN CORP.
 PA Shinkens RA, Leach M;
 PI WPI, 2000-602362/57.
 DR N-PSDB; AAC76790.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX Claim 11; Page 3880-3881; 5507pp; English.
 XX AACT74446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cyrostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; nootropic; neuroprotective; osteopathic;
 CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;
 CC cardiatic; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antihemetic; antihydrolytic; and antianemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX SQ Sequence 299 AA;
 Alignment Scores:
 Pred. No.: 7,15e-67 Length: 299
 Score: 877.00 Matches: 174
 Percent Similarity: 98.3% Conservative: 0
 Best Local Similarity: 98.3% Mismatches: 3
 Query Match: 83.6% Indels: 0
 DB: 3 Gaps: 0
 US-10-699-035A-1 (1-537) x AAB42581 (1-299)
 QY 4 GACCTGATGTTCTGCTGACAGCTCAGCCAGCGTCTTCACTACAGTTTCCCGGTT 63
 DB 37 AaPLeuMeLpHeuLeuAaPSeRSeRaLaSeRValSeRHiStyLcLpHeSeRaTgVal 56
 QY 64 CGGAGCTTGTGGGGAGCTGGGCTGACATGCCCCCTGGGACCGGGGCTTGGTGGC 123
 DB 57 ATGgLuPheValGlgInLeuValAlaProLeuProLeuAlaPro**AlaLeuAArgAla 76

QY 124 AGCTGTGTCACGTGGGCGAGTGGCCATACCGAGTTCCCTTGGGCCACACAGCTCG 183
 DB 77 SerLeuValHiStValGlySeRaRProTyThrGluPheProPheGlyGlnHiSeSeR 96
 QY 184 GGTAGGCTGGCCAGAGATGCGGCTTCCTTCGCCAGCGGCATGGGTGACACCACT 243
 DB 97 GlyGluAlaAlaGlnAaPSeRaLaValaTgAlaSeRLaGlnAaTgMetGlyAaPThRHiSeThr 116
 QY 244 GGCCTGGCGCTGTGCTAAGCAAGAAACAGCTGTTTGTGAAGCATCAGTGGCCGCCA 303
 DB 117 GlyLeuAlaLeuValTYRAlaLysGluGlnLeuHeaLaGluAlaSeRGLAlaArgPro 136
 QY 304 GGGGTGCCCAAGTGTCTGTGTGAGTACAGATGGCGGCTCCAGGACCTGTGGGCC 363
 DB 137 GlyValProLysValLeuValITrpValIThrAaPglGlySeRSeRaPProValGlyPro 156
 QY 364 CCCATGACAGAGCTCAAGAGACCTGGCGCTCACCGTTCATTGTCAGACCGGCGAGGC 423
 DB 157 ProMetGlnGluLeuLysAaPLeuGlyValIThrValIPheIleValSeRThRGLAArgGly 176
 QY 424 AACTTCTGAGCTGTGACCGGCTGCTGACCGCCCTGCGGAGAGACCTGCATTGTG 483
 DB 177 AaPHeuLeuGluLeuSeRaLaAlaLaSeRaLaProAlaGluLysHiSeuHiPheVal 196
 QY 484 GACGTGATGACCTGCACATCATTTGTCCAGAGCTGAGGGGCTCCATTCTC 534
 DB 197 AaPValAaPSeRaPLeuHiStIleValGlnGlnLeuAaTgGlySeRILEu 213
 RESULT 14
 AAE32503
 ID AAE32503 standard; protein, 180 AA.
 XX AAE32503;
 AC 24-MAR-2003 (first entry)
 XX Mouse von Willebrand Factor A (VA) domain.
 DE Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
 KM extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
 KM gene therapy; mouse.
 OS Mus sp.
 OS Key Location/Qualifiers
 FH Misc-difference 1..2
 FT "Encoded by GGG"
 FT WO200288184-A1.
 PN 07-NOV-2002.
 XX 02-MAY-2002; 2002WO-AU000542.
 PF 02-MAY-2001; 2001AU-00004701.
 XX (MURDOCH CHILDRENS RES INST.
 PA Bateman JF, Fitzgerald DJ;
 PI WPI, 2003-111873/10.
 DR N-PSDB; AAD50400.
 XX New isolated Willebrand Factor A-Related Protein polypeptide useful for
 PT the manufacture of a medicament in the treatment of a disease condition
 PT of the extracellular matrix, in particular arthritis.
 PS Claim 9; Page 78-79; 103pp; English.
 CC The invention relates to Willebrand Factor A domain related-protein
 CC (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
 CC superfamily of extracellular matrix (ECM) proteins. WARP is used as a

CC molecular marker, used for detecting a loss of ECM integrity in an animal
CC subject, monitoring repair, regeneration or other disease processes in an
CC animal subject and detecting a disease condition or a propensity for the
CC development of a disease condition in an animal subject. The invention is
CC useful for the manufacture of a medication in the treatment of a disease
CC condition of the ECM. The disease condition involves the cartilage, and
CC is preferably arthritis. The invention is also used in gene therapy. The
CC present sequence is mouse VA domain

SQ Sequence 180 AA;

Alignment Scores:

Pred. No.:	5,31e-57	Length:	180
Score:	762.00	Matches:	149
Percent Similarity:	91.5%	Conservative:	13
Best Local Similarity:	84.2%	Mismatches:	15
Query Match:	72.6%	Indels:	0
DB:	6	Gaps:	0

US-10-699-035A-1 (1-537) x AAE32503 (1-180)

QY 1 GGGGACCTGATGTCCTGCTGACAGCTGACGAGGCTCTCTCACTACGAGTTCTCCCG 60
DB 2 GlyAspLeuLeuPheLeuLeuAspSerAlaSerValSerHisTyrGluPheSerArg 21
QY 61 GTTCGGAGATTGTGGGGAGAGCTGTGCTGCTCACTGCCCCCTGGAGCCGCGCT 120
DB 22 ValArgGluPheValGlyGlnLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 41
QY 121 GCCAGTCTGTGACAGCTGGGACAGTCCGACCATACAGGAGTCCCTTGGCCAGCAGC 180
DB 42 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 61
QY 181 TCGGGTGAAGCTCCAGAGTGGGCTGGCGCTTCTGCCAGGACGAGGTGACACCCAC 240
DB 62 SerGlyGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspTrpAsn 81
QY 241 ACTGCGCTGGCGCTGTGCTATGCCAGAAACAGCTGTTGCTAGACATCAGGTGCCCG 300
DB 82 ThrGlyLeuAlaLeuAlaTyrAlaTyrGluGlnLeuPheLeuGlnGluAlaGlyAlaArg 101
QY 301 CCAGGGGGTCCCAAAGTGTGCTGTGGTGTGACAGATGGCGGCTCCAGCCAGCTGTGGC 360
DB 102 ProGlyValProGlyValLeuValITPValIThAspGlyGlySerSerAspProValGly 121
QY 361 CCCCCATGAGGAGCTCAAGAGCTGGGGCTGACCGGTTCATTGTTCAGCACCGGCGCA 420
DB 122 ProPheMetGlnGlnLeuLeuYsaPheGlyValIThrIlePheIleValSerIThrGlyArg 141
QY 421 GGCACCTTCTGAGCTGTGACCGGCTGCTCAAGCCCTGCCAGAACAGCACTGCACTT 480
DB 142 GlyAsrLeuLeuGlnLeuLeuAlaAlaIleAsrAlaProIleGlnIleHisLeuHisPhe 161
QY 481 GTGACGTGATGACCTGACATCATTTGTCAAAGACTGAGGGGCTTCATT 531
DB 162 ValAspValAspAspLeuProIleIleAlaArgGluLeuArgGlySerIle 178

RESULT 15

AAE32501

ID AAE32501 standard; protein; 415 AA.

XX AAE32501;

DT 24-MAR-2003 (first entry)

DE Mouse Willebrand Factor A domain related-protein (WARP).

XX Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;

XX extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;

XX gene therapy; mouse.

XX Mus sp.

FN Key Location/Qualifiers
FT Peptide 1..18
FT Protein /label= Signal_peptide
FT Protein /note= "Mouse mature WARP protein"
FT Modified-site 148
FT Modified-site /note= "O-glycosylation site"
FT Modified-site 264
FT Modified-site /note= "N-glycosylation site"
FT Modified-site 359
FT Modified-site /note= "N-glycosylation site"
FT Modified-site 361
FT Disulfide-bond /note= "O-glycosylation site"
FT Modified-site 369..393
FT Modified-site /note= "O-glycosylation site"
PN MO200288184-A1.
PD 07-NOV-2002.
PF 02-MAY-2002; 2002MO-AU000542.
PK 02-MAY-2001; 2001AU-00004701.
XX (MURD-) MURDOCH CHILDRENS RES INST.
XX Bateman JF, Fitzgerald DJ;
XX WPI, 2003-111873/10.
XX N-PSDB; AAD50022, AAD50398.
PT New isolated Willebrand Factor A-Related Protein polypeptide useful for
PT the manufacture of a medication in the treatment of a disease condition
PT of the extracellular matrix, in particular arthritis.
XX Claim 13; Page 74-75; 103pp; English.
XX The invention relates to Willebrand Factor A domain related-protein
XX (WARP) which is a member of von Willebrand Factor A (vW)-domain protein
XX superfamily of extracellular matrix (ECM) proteins. WARP is used as a
XX molecular marker, used for detecting a loss of ECM integrity in an animal
XX subject, monitoring repair, regeneration or other disease processes in an
XX animal subject and detecting a disease condition or a propensity for the
XX development of a disease condition in an animal subject. The invention is
XX useful for the manufacture of a medication in the treatment of a disease
XX condition of the ECM. The disease condition involves the cartilage, and
XX is preferably arthritis. The invention is also used in gene therapy. The
XX present sequence is mouse WARP protein

SQ Sequence 415 AA;

Alignment Scores:

Pred. No.:	6,28e-57	Length:	415
Score:	762.00	Matches:	149
Percent Similarity:	91.5%	Conservative:	13
Best Local Similarity:	84.2%	Mismatches:	15
Query Match:	72.6%	Indels:	0
DB:	6	Gaps:	0

US-10-699-035A-1 (1-537) x AAE32501 (1-415)

QY 1 GGGGACCTGATGTCCTGCTGACAGCTGACGAGGCTCTCTCACTACGAGTTCTCCCG 60
DB 33 GlyAspLeuLeuPheLeuLeuAspSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCGGAGATTGTGGGGAGAGCTGTGCTGCTCACTGCCCCCTGGAGCCGCGCT 120
DB 53 ValArgGluPheValGlyGlnLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 72
QY 121 GCCAGTCTGTGACAGCTGGGACAGTCCGACCATACAGGAGTCCCTTGGCCAGCAGC 180
DB 73 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 92

```

QY 181 TCGGGTGAAGCTGCCAGAGATGCGGTGCGTCTTCTGCCAGCGCATGCGTGACACCCAC 240
Db 93 SerGlyInAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 112
QY 241 ACTGCGCTGGCGCGCTGTCTATGCCAAGAACACAGCTGTTTGTGAAGCATCAGGTGCCGG 300
Db 113 ThrGlyLeuAlaLeuAlaTyrAlaLysGlnLeuPheAlaGlnGluAlaGlyAlaArg 132
QY 301 CCAGGGGTGCCCAAGTGTGTGTGGTGACAGATGGCGGCTCCAGCGACCCCTGTGGGC 360
Db 133 ProGlyValProLysValLeuValTrrPvalThrAspGlyGlySerSerAspProValGly 152
QY 361 CCCCCATGCAAGAGCTCAAGAGACCTGGGCGTCAACGGTGTTCATTGTGACACCGGCCGA 420
Db 153 ProPrometGlnGlnLeuLysAspLeuGlyValThrIlePheIleValSerThrGlyArg 172
QY 421 GGCACTTCCTGGAGACTGTCAAGCCGCTGCTCAGCCCTCGCGAGAGAAGACCTGCACCTT 480
Db 173 GlyAsnLeuLeuGlnLeuLeuAlaAlaSerAlaProAlaGlnLysHisLeuHisPhe 192
QY 481 GTGACGTGATGATGATGATGATCATTTGCAAGAGAGTGAAGGGCTCCATT 531
Db 193 ValAspValAspAspLeuProIleIleAlaArgGlnLeuArgGlySerIle 209

```

Search completed: February 13, 2006, 13:23:30
 Job time : 97.9481 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_r2p model

Run on: February 13, 2006, 13:42:00 ; Search time 17.5402 Seconds
(without alignments)
2558.401 Million cell updates/sec

Title: US-10-699-035A-1
Perfect score: 1049
Sequence: 1 gggagaccgacgtctcctcgtc.....tgaagggtccatctcgcg 537

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1867569 seqs, 417829326 residues
Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ r2p model -DEV=xld
-O=/abes/ABSWEB_spool/US10699035/runat_13022006_062450_25572/app_query.fasta_1
-DB=Published Applications_AA_Main-OPMT=fastan-SUFFIX=rapbm-MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abs02p
-USER=US10699035@cgn1.1.405@runat_13022006_062450_25572 -NCPU=6 -ICPU=3
-NO MAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA_Main:*

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep:*
- 3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*
- 4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	902	86.0	180	4	US-10-699-035A-2
2	902	86.0	215	3	US-09-789-561-85
3	902	86.0	215	3	US-09-833-245-2096
4	902	86.0	215	5	US-10-883-936-85
5	902	86.0	226	3	US-09-789-561-165
6	902	86.0	226	5	US-10-883-936-165
7	902	86.0	242	3	US-09-789-561-159
8	902	86.0	242	5	US-10-883-936-159
9	902	86.0	418	4	US-10-699-035A-6
10	902	86.0	418	4	US-10-699-035A-20
11	886	84.5	185	4	US-10-149-819-18

12	762	72.6	180	4	US-10-699-035A-8	Sequence 8, Appli
13	762	72.6	181	4	US-10-699-035A-31	Sequence 31, Appli
14	762	72.6	415	4	US-10-699-035A-4	Sequence 4, Appli
15	762	72.6	415	4	US-10-699-035A-21	Sequence 21, Appli
16	751	71.6	186	3	US-09-789-561-160	Sequence 160, App
17	751	71.6	186	5	US-10-883-936-160	Sequence 160, App
18	727.5	26.0	176	3	US-09-976-782-38	Sequence 38, Appli
19	265.5	25.3	182	4	US-10-699-035A-22	Sequence 22, Appli
20	263.5	25.1	184	4	US-10-699-035A-27	Sequence 27, Appli
21	261.5	24.9	3063	3	US-09-918-715-257	Sequence 257, App
22	261.5	24.9	3063	4	US-10-177-293-61	Sequence 61, Appli
23	261.5	24.9	3063	4	US-10-177-293-61	Sequence 63, Appli
24	261.5	24.9	3063	4	US-10-301-822-26	Sequence 26, Appli
25	261.5	24.9	3063	4	US-10-474-794-257	Sequence 257, App
26	261.5	24.9	3063	5	US-10-979-159-257	Sequence 257, App
27	261.5	24.9	3063	5	US-10-631-467-918	Sequence 918, App
28	261.5	24.9	3118	5	US-10-220-335-287	Sequence 287, App
29	259	24.7	580	5	US-10-723-860-2769	Sequence 2769, Ap
30	254.5	24.3	755	3	US-09-919-497-57	Sequence 57, Appli
31	254.5	24.3	1207	4	US-10-408-765A-1591	Sequence 1591, Ap
32	254.5	24.3	1297	4	US-10-187-975-102	Sequence 102, App
33	254.5	24.3	1780	4	US-10-115-479-10	Sequence 10, Appli
34	247.5	23.6	415	3	US-09-976-782-114	Sequence 114, App
35	247.5	23.6	493	3	US-09-976-782-115	Sequence 115, App
36	246.5	23.5	776	4	US-10-000-512-8	Sequence 8, Appli
37	246.5	23.5	776	4	US-10-074-566-8	Sequence 176, Appli
38	246.5	23.5	782	4	US-10-428-275-176	Sequence 1292, Ap
39	246.5	23.5	794	3	US-09-833-245-1292	Sequence 16, Appli
40	246.5	23.5	896	4	US-10-004-378A-16	Sequence 158, App
41	246.5	23.5	896	4	US-10-428-275-158	Sequence 156, App
42	246.5	23.5	914	4	US-10-428-275-156	Sequence 34, Appli
43	246.5	23.5	915	3	US-09-909-320-34	Sequence 34, Appli
44	246.5	23.5	915	3	US-09-909-088B-34	Sequence 34, Appli
45	246.5	23.5	915	3	US-09-905-291A-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1
US-10-699-035A-2
; Sequence 2, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-2

Alignment Scores:

Pred. No.: 7.36e-65
Score: 902.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 86.0%
DB: 4
Length: 180
Matches: 178
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-2 (1-180)

QY 1 GGGACCTGATGTTCTCTGCGACGCTGACGACGCTTCTTCACTAGAGTTTCCCGG 60
|||||

Db 2 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 21
QY GTTCGGAGATTGTGGGGCAGCTGTGGCTCCACTGCGCCCTGGGACCGGGGCTTGAGT 120
Db 22 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 41
QY 121 GCCAGTCTGTGCACGTGGGAGTCGGCCATACACCGAGTTCCCTTGGCCAGACAGC 180
Db 42 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 61
QY 181 TCGGGTGAAGCTGCCAGAGATGCCGTGCTTTCGCCAGCGAGATGGTGAACACCCAC 240
Db 62 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 81
QY 241 ACTGGCTGTGGCGCTGTATGCTCAAGGAACAGCTGTTGTGTAAGCATCAGGTGCCG 300
Db 82 ThrGlyLeuAlaLeuValTyrAlaGlyGlnLeuPheAlaGlnHisSerGlyAlaArg 101
QY 301 CCAGGGGTGCCCAAGTGTGTGTGGTGCACAGATGGCGGCTCCAGCGACCTGTGGGC 360
Db 102 ProGlyValProGlyValLeuValTyrValThrAspGlyGlySerSerAspProValGly 121
QY 361 CCCCCATGCAGAGACTCAAGACCTGGCGCTCACCGTGTTCATTGTGCACGCGCGCA 420
Db 122 ProPrometGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 141
QY 421 GGCAACTTCCTGGAGCTGTCAAGCCGCTGCGCTCAGCCCTCGAGAGACACTGCATT 480
Db 142 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisSerPhe 161
QY 481 GTGACGTGATGACCTGCACATCTGTGCCAAGCTGAGGGGCTCCATTCTC 534
Db 162 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 179

RESULT 2

US-09-789-561-85
; Sequence 85, Application US/09789561
; Patent No. US2002064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-85

Alignment Scores:

Pred. No.: 7,52e-65 Length: 215
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 3 Gaps: 0

US-10-699-035A-1 (1-537) x US-09-789-561-85 (1-215)

QY 1 GGGGACCTGATGTTCTCTGCTGACAGCTCAGCCAGCGTCTTCACACTAGAGTTTCCCGG 60

Db 33 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
QY 61 GTTCGGAGATTGTGGGGCAGCTGTGGCTCCACTGCGCCCTGGGACCGGGGCTTGAGT 120
Db 53 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
QY 121 GCCAGTCTGTGCACGTGGGAGTCGGCCATACACCGAGTTCCCTTGGCCAGACAGC 180
Db 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
QY 181 TCGGGTGAAGCTGCCAGAGATGCCGTGCTTTCGCCAGCGCATGGGTGAACACCCAC 240
Db 93 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 112
QY 241 ACTGGCTGTGGCGCTGTATGCTCAAGGAACAGCTGTTGTGTAAGCATCAGGTGCCG 300
Db 113 ThrGlyLeuAlaLeuValTyrAlaGlyGlnLeuPheAlaGlnHisSerGlyAlaArg 132
QY 301 CCAGGGGTGCCCAAGTGTGTGTGGTGCACAGATGGCGGCTCCAGCGACCTGTGGGC 360
Db 133 ProGlyValProGlyValLeuValTyrValThrAspGlyGlySerSerAspProValGly 152
QY 361 CCCCCATGCAGAGACTCAAGACCTGGCGCTCACCGTGTTCATTGTGCACGCGCGCA 420
Db 153 ProPrometGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 172
QY 421 GGCAACTTCCTGGAGCTGTCAAGCCGCTGCGCTCAGCCCTCGAGAGACACTGCATT 480
Db 173 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisSerPhe 192
QY 481 GTGACGTGATGACCTGCACATCTGTGCCAAGCTGAGGGGCTCCATTCTC 534
Db 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 210

RESULT 3

US-09-833-245-2096
; Sequence 2096, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PFS46PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2096
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2096

Alignment Scores:

Pred. No.: 7,52e-65 Length: 215
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 3 Gaps: 0

US-10-699-035A-1 (1-537) x US-09-833-245-2096 (1-215)

QY 1 GGGACCTGATGTTCTCTGCTGGACAGCTAGCCAGGCTTCTCACTACGAGTTTCCCGG 60
 |||||
 Db 33 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
 |||||
 QY 61 GTTCGGAGATTGTGTGGGAGAGCTGGTGCTCCACTGCCCCCTGGAGCCCGGCGCT 120
 |||||
 Db 53 ValArgGluPheValGlyGluLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
 |||||
 QY 121 GCCAGTCTGTGTCACAGTGGGAGCTGGCCCATACACCGAGTTCCCTTCCGAGCAGAC 180
 |||||
 Db 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
 |||||
 QY 181 TCGGGTGAAGCTGCCCAAGATGGCGGTGCTCTTCTGCCAGGCACTGGGTACACCCAC 240
 |||||
 Db 93 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 112
 |||||
 QY 241 ACTGGCTGGCGCTGGTCTATGCCCAAGAGACAGCTGTTTGTGAAGCATCAGTCCCGG 300
 |||||
 Db 113 ThrGlyLeuAlaLeuValTyrAlaGlyGlnLeuPheAlaGluAlaSerGlyAlaArg 132
 |||||
 QY 301 CCAGGGGTGCCCAAGTGTCTGTGGTGAGACAGATGGCGGCTCCAGCAGCCCTGTGGGC 360
 |||||
 Db 133 ProGlyValProLysValLeuValTyrValThrAspGlyGlySerSerAspProValGly 152
 |||||
 QY 361 CCCCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCTGTTTCATTGTCAGACCGGCGCA 420
 |||||
 Db 153 ProPheMetGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 172
 |||||
 QY 421 GGCACTTCTGTGAGCTGTACCGCTGCTGCTCAGCCCCCTGGAGAGAGACCTGCACTTT 480
 |||||
 Db 173 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 192
 |||||
 QY 481 GTGACGTGATGACCTGACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
 |||||
 Db 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 210
 |||||

RESULT 4

US-10-883-936-85
 ; Sequence 85, Application US/10883936
 ; Publication No. US20050019866A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NI et al.
 ; TITLE OF INVENTION: 52 Human secreted proteins
 ; FILE REFERENCE: P2043P1
 ; CURRENT APPLICATION NUMBER: US/10/883,936
 ; PRIORITY FILING DATE: 2004-07-06
 ; PRIOR APPLICATION NUMBER: US/09/789,561
 ; PRIORITY FILING DATE: 2001-02-22
 ; PRIOR APPLICATION NUMBER: PCT/US00/24008
 ; PRIOR FILING DATE: 2000-08-31
 ; PRIOR APPLICATION NUMBER: 60/152,317
 ; PRIOR FILING DATE: 1999-09-03
 ; PRIOR APPLICATION NUMBER: 60/152,315
 ; PRIOR FILING DATE: 1999-09-03
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 85
 ; LENGTH: 215
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (7)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-10-883-936-85

Alignment Scores:

Pred. No.: 7,52e-65 Length: 215
 Score: 902.00 Matches: 178
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 86.0% Indels: 0
 DB: 5 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-883-936-85 (1-215)
 QY 1 GGGACCTGATGTTCTCTGCTGGACAGCTAGCCAGGCTTCTCACTACGAGTTTCCCGG 60
 |||||
 Db 33 GlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
 |||||
 QY 61 GTTCGGAGATTGTGTGGGAGAGCTGGTGCTCCACTGCCCCCTGGAGCCCGGCGCT 120
 |||||
 Db 53 ValArgGluPheValGlyGluLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 72
 |||||
 QY 121 GCCAGTCTGTGTCACAGTGGGAGCTGGCCCATACACCGAGTTCCCTTCCGAGCAGAC 180
 |||||
 Db 73 AlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 92
 |||||
 QY 181 TCGGGTGAAGCTGCCCAAGATGGCGGTGCTCTTCTGCCAGGCACTGGGTACACCCAC 240
 |||||
 Db 93 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 112
 |||||
 QY 241 ACTGGCTGGCGCTGGTCTATGCCCAAGAGACAGCTGTTTGTGAAGCATCAGTCCCGG 300
 |||||
 Db 113 ThrGlyLeuAlaLeuValTyrAlaGlyGlnLeuPheAlaGluAlaSerGlyAlaArg 132
 |||||
 QY 301 CCAGGGGTGCCCAAGTGTCTGTGGTGAGACAGATGGCGGCTCCAGCAGCCCTGTGGGC 360
 |||||
 Db 133 ProGlyValProLysValLeuValTyrValThrAspGlyGlySerSerAspProValGly 152
 |||||
 QY 361 CCCCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCTGTTTCATTGTCAGACCGGCGCA 420
 |||||
 Db 153 ProPheMetGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 172
 |||||
 QY 421 GGCACTTCTGTGAGCTGTACCGCTGCTGCTCAGCCCCCTGGAGAGAGACCTGCACTTT 480
 |||||
 Db 173 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 192
 |||||
 QY 481 GTGACGTGATGACCTGACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
 |||||
 Db 193 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 210
 |||||

RESULT 5

US-09-789-561-165
 ; Sequence 165, Application US/09789561
 ; Patent No. US20020064818A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NI et al.
 ; TITLE OF INVENTION: 52 Human secreted proteins
 ; FILE REFERENCE: P2043P1
 ; CURRENT APPLICATION NUMBER: US/09/789,561
 ; PRIORITY FILING DATE: 2001-02-22
 ; PRIOR APPLICATION NUMBER: PCT/US00/24008
 ; PRIOR FILING DATE: 2000-08-31
 ; PRIOR APPLICATION NUMBER: 60/152,317
 ; PRIOR FILING DATE: 1999-09-03
 ; PRIOR APPLICATION NUMBER: 60/152,315
 ; PRIOR FILING DATE: 1999-09-03
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 165
 ; LENGTH: 226
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-789-561-165

Alignment Scores:

Pred. No.: 7,57e-65 Length: 226
 Score: 902.00 Matches: 178
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 86.0% Indels: 0
 DB: 3 Gaps: 0

US-10-699-035A-1 (1-537) x US-09-789-561-165 (1-226)

QY 1 GGGGACCTGATGTTCTCTGCTGAGACAGCTCAGCCAGCTCTCTCACTACGAGTTCTCCCGG 60
Db 46 GlyAspLeuMetPheLeuLeuAspSerSerLaseValSerHisTyrGluPheSerArg 65
QY 61 GTTGGGAGTTTGTGGGAGAGCTGATGCTCAGTCCCGGAGCCGAGGAGCCCTGCGT 120
Db 66 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 85
QY 121 GCCAGTGTGTGACAGTGGGAGTGGCCATACACCGAGTTCCCTTCGGCCAGCACAGC 180
Db 86 AlaserLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 105
QY 181 TCGGATGAGGCTGCCCGAGAGATGCGGTGCTTTCGCCAGCGATGGGTGACACCCAC 240
Db 106 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 125
QY 241 ACTGGCTGGCGCTGCTATATGCCAAAGAACAGCTGTTGCTGAAGCATCAGTGGCCCGG 300
Db 126 ThrGlyLeuAlaLeuValTyrAlaIlySerGluGlnLeuPheAlaGlnAlaSerGlyAlaArg 145
QY 301 CCAAGGAGTGGCCAAAGTGTGTGTGGTGAACATGGCGGCTCCAGCAGCCTGTGGGC 360
Db 146 ProGlyValProIysValLeuValTyrValThrAspGlyGlySerSerAspProValGly 165
QY 361 CCCCCATGACAGAGCTCAAGAGCTGGGCGCTCAGCGTTCATTGTACAGACCGCGCGA 420
Db 166 ProPheMetGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 185
QY 421 GGCAACTTCCTGGAGAGCTGTACAGCGCTGCTCAGCGCCGCGAGAGACCTGCACATT 480
Db 186 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 205
QY 481 GTGACGTGTGATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
Db 206 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 223

RESULT 6

US-10-883-936-165
; Sequence 165, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 165
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-883-936-165

Alignment Scores:

Pred. No.: 7.57e-65 Length: 226
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 5 Gaps: 0

US-10-699-035a-1 (1-537) x US-10-883-936-165 (1-226)

QY 1 GGGGACCTGATGTTCTCTGCTGAGACAGCTCAGCCAGCTCTCTCACTACGAGTTCTCCCGG 60

Db 46 GlyAspLeuMetPheLeuLeuAspSerSerLaseValSerHisTyrGluPheSerArg 65
QY 61 GTTGGGAGTTTGTGGGAGAGCTGATGCTCAGTCCCGGAGCCGAGGAGCCCTGCGT 120
Db 66 ValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArg 85
QY 121 GCCAGTGTGTGACAGTGGGAGTGGCCATACACCGAGTTCCCTTCGGCCAGCACAGC 180
Db 86 AlaserLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSer 105
QY 181 TCGGATGAGGCTGCCCGAGAGATGCGGTGCTTTCGCCAGCGATGGGTGACACCCAC 240
Db 106 SerGlyGluAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHis 125
QY 241 ACTGGCTGGCGCTGCTATATGCCAAAGAACAGCTGTTGCTGAAGCATCAGTGGCCCGG 300
Db 126 ThrGlyLeuAlaLeuValTyrAlaIlySerGluGlnLeuPheAlaGlnAlaSerGlyAlaArg 145
QY 301 CCAAGGAGTGGCCAAAGTGTGTGTGGTGAACATGGCGGCTCCAGCAGCCTGTGGGC 360
Db 146 ProGlyValProIysValLeuValTyrValThrAspGlyGlySerSerAspProValGly 165
QY 361 CCCCCATGACAGAGCTCAAGAGCTGGGCGCTCAGCGTTCATTGTACAGACCGCGCGA 420
Db 166 ProPheMetGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArg 185
QY 421 GGCAACTTCCTGGAGAGCTGTACAGCGCTGCTCAGCGCCGCGAGAGACCTGCACATT 480
Db 186 GlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHisPhe 205
QY 481 GTGACGTGTGATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
Db 206 ValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 223

RESULT 7

US-09-789-561-159
; Sequence 159, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 159
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-159

Alignment Scores:

Pred. No.: 7.63e-65 Length: 242
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 3 Gaps: 0

US-10-699-035a-1 (1-537) x US-09-789-561-159 (1-242)

```
QY 1 GGGGACCTGATGTTCTGCTGGAACAGCTCAGCCAGCGTCTCTACTAGAGTTCTCCGG 60
DB 60 GYAAspleuMet.PheleuLeuAspSerSerAlaSerValSerHisIerIcIupheserArg 79
QY 61 GTTCCGAGATTGTTGGGAGAGCTGCGTCCAGTCCCGCTGGGACCGGGCCCTGGCT 120
DB 80 VALArgIupheValGIyGInleuValAlaProleuProleuGIyThrcIyAlaLeuArg 99
QY 121 GCCAGTCTGTCACAGTGGGAGTCCGATACACCGATTCCCTTCGGCCAGACAGC 180
DB 100 AlaSerLeuValHisValGIySerArgProIyThrcIupheProheGIyGInHisSer 119
QY 181 TCGGGTAGCGTCCCGCAGATGCGGTGCTTCTGCCAGCGCATGGGTGACACCCAC 240
DB 120 SerGIyGIuAlaIaGIInAspAlaValArgAlaSerAlaGIInArgMetGIyAspThrHis 139
QY 241 ACTGCGCTGGCGCTGCTCTATGCGCAAGAAAGCTGTTGTCAGACATCAGTCCCGG 300
DB 140 ThrGIyLeuAlaLeuValIyAlaIySgIuGInleuPheAlaGIuAlaSerGIyAlaArg 159
QY 301 CCAGGGAGTCCCAAAGTGTGCTGTTGGTGAACAGATGCGGCTCCAGCAGCCTGGAGC 360
DB 160 ProGIyValProIyValIeValIyThrcIupheProIyGIySerSerAspProValGIy 179
QY 361 CCCCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCGTTCATTGTGACACCGCCGA 420
DB 180 ProprometGIInGIuLeuIySAspLeuGIyValIThrcIupheIleValSerThrGIyArg 199
QY 421 GCGCAATTCCTGAGAGCTGTCAGCGCTGCTCAGCCCTCCGAGAGACATGACCTTT 480
DB 200 GYAAspPheLeuGIuLeuSerAlaAlaIaSerAlaProIaGIuIySHisIeunHisPhe 219
QY 481 GTGACGTGATGACCTGCACATCATTTGTCCAAGAGCTGAGGGGCTCCATTCTC 534
DB 220 VALAspValAspAspLeuHisIleIleValGIInGIuLeuArgGIySerIleLeu 237

RESULT 8
US-10-883-936-159
; Sequence 159, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: N1 et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 159
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
```

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-159

Alignment Scores:
Pred. No.: 7,63e-65 Length: 242
Score: 902.00 Matches: 178
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 86.0% Indels: 0
DB: 5 Gaps: 0

US-10-699-035a-1 (1-537) x US-10-883-936-159 (1-242)

```
QY 1 GGGGACCTGATGTTCTGCTGGAACAGCTCAGCCAGCGTCTCTACTAGAGTTCTCCGG 60
DB 60 GYAAspleuMet.PheleuLeuAspSerSerAlaSerValSerHisIerIcIupheserArg 79
QY 61 GTTCCGAGATTGTTGGGAGAGCTGCGTCCAGTCCCGCTGGGACCGGGCCCTGGCT 120
DB 80 VALArgIupheValGIyGInleuValAlaProleuProleuGIyThrcIyAlaLeuArg 99
QY 121 GCCAGTCTGTCACAGTGGGAGTCCGATACACCGATTCCCTTCGGCCAGACAGC 180
DB 100 AlaSerLeuValHisValGIySerArgProIyThrcIupheProheGIyGInHisSer 119
QY 181 TCGGGTAGCGTCCCGCAGATGCGGTGCTTCTGCCAGCGCATGGGTGACACCCAC 240
DB 120 SerGIyGIuAlaIaGIInAspAlaValArgAlaSerAlaGIInArgMetGIyAspThrHis 139
QY 241 ACTGCGCTGGCGCTGCTCTATGCGCAAGAAAGCTGTTGTCAGACATCAGTCCCGG 300
DB 140 ThrGIyLeuAlaLeuValIyAlaIySgIuGInleuPheAlaGIuAlaSerGIyAlaArg 159
QY 301 CCAGGGAGTCCCAAAGTGTGCTGTTGGTGAACAGATGCGGCTCCAGCAGCCTGGAGC 360
DB 160 ProGIyValProIyValIeValIyThrcIupheProIyGIySerSerAspProValGIy 179
QY 361 CCCCCCATGACAGAGCTCAAGAGCTGGGCGTCAACCGTTCATTGTGACACCGCCGA 420
DB 180 ProprometGIInGIuLeuIySAspLeuGIyValIThrcIupheIleValSerThrGIyArg 199
QY 421 GCGCAATTCCTGAGAGCTGTCAGCGCTGCTCAGCCCTCCGAGAGACATGACCTTT 480
DB 200 GYAAspPheLeuGIuLeuSerAlaAlaIaSerAlaProIaGIuIySHisIeunHisPhe 219
QY 481 GTGACGTGATGACCTGCACATCATTTGTCCAAGAGCTGAGGGGCTCCATTCTC 534
DB 220 VALAspValAspAspLeuHisIleIleValGIInGIuLeuArgGIySerIleLeu 237

RESULT 9
US-10-699-035a-6
; Sequence 6, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 418
; TYPE: PRT
```



```

FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030044913A1 3683905CD1
US-10-149-819-18

Alignment Scores:
Pred. No.: 1,486-63 Length: 185
Score: 886.00 Matches: 175
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 84.5% Indels: 0
DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-149-819-18 (1-185)

QY 10 ATGTTCTCTGAGACAGCTCAGCCAGCGCTCTCACTACAGATTCTCCGGGTTCCGAG 69
DB 1 MetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArgValaGlu 20
QY 70 TTGTGGGGCAGCTGTGTGCTCCTCACTGCGCCCTGGGACCGGGGCGCTGCGCACTCTG 129
DB 21 PheValGlyGluLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeu 40
QY 130 GTGCACGTGGGCACTGGCCATACACCGAGTTCCCTTGGCCAGCACAGCTCGGGTGA 189
DB 41 ValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGlu 60
QY 190 GGTGCCAGAGATGGCGGTGGCTTGTGCGCCAGCGCATGGGAGACCAACCACTGGCGTG 249
DB 61 AlaAlaGlnAspAlaValAlaArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeu 80
QY 250 GCGCTGTCTATGCCAAGAACAGCTGTTGTCTGAACATCAGTGTCCCGGACAGGGGTG 309
DB 81 AlaLeuValTyrAlaIleGluGlnLeuPheAlaGluAlaSerCylAlaArgProGlyVal 100
QY 310 CCCAAAGTCTGTGTGTGTGACAGATGGCGGTCTCCAGGACCCCTTGGGCGCCCGCAAG 369
DB 101 ProLysValLeuValTyrValThrAspGlyGlySerSerAspProValGlyProPomMet 120
QY 370 CAGGAGCTCAAGGACCTGGGCGGTCAACGTTTCATTTGATGAGACCGGCGGCAAGCTTC 429
DB 121 GlnGluLeuLeuAspLeuLeuValThrValPheIleValSerThrGlyArgGlyAsnPro 140
QY 430 CTGGAAGCTGTCAAGCCGCTGCTCAGCCCTGCGGAGAACACCTGCACTTTGTGACGTG 489
DB 141 LeuGluLeuSerAlaAlaIleSerAlaProAlaGluLeuHisLeuHisPheValaAspVal 160
QY 490 GATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATTCTC 534
DB 161 AspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 175

RESULT 12
US-10-699-035A-8
Sequence 8, Application US/10699035A
Publication No. US20040214349A1
GENERAL INFORMATION:
APPLICANT: Bateman, John
TITLE OF INVENTION: A Molecular Marker
FILE REFERENCE: A36056 PCT USA A 071838.0142
CURRENT APPLICATION NUMBER: US/10/699,035A
PRIOR FILING DATE: 2003-10-31
PRIOR APPLICATION NUMBER: PCT/AU02/00542
PRIOR FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: AU PR4701/01
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 180
TYPE: PRT
ORGANISM: Mus musculus
US-10-699-035A-8

```

```

Alignment Scores:
Pred. No.: 1,836-53 Length: 180
Score: 762.00 Matches: 149
Percent Similarity: 91.5% Conservative: 13
Best Local Similarity: 84.2% Mismatches: 15
Query Match: 72.6% Indels: 0
DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-8 (1-180)

QY 1 GGGGACCTGATGTTCTCTGTCAGCTCAGCCAGCGTCTCACTACAGATTCTCCGGG 60
DB 2 GlyAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 21
QY 61 GTTCCGAGATTGTGTGGGACAGCTGTGTGCTCCACCTGCGGACCGGGGCGCTGCGCT 120
DB 22 ValArgGluPheValGlyGlnLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 41
QY 121 GCCACTGTGTGCACAGTGGGACAGTGGCCATACACCGAGTTCCCTTGGCCAGCACAGC 180
DB 42 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 61
QY 181 TTGGGTGAGGCTGCGCAGATGCGGTGCTTGTGCCCGACGATGGGTACACCCAC 240
DB 62 SerGlyGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 81
QY 241 ACTGGCCTGGGCGCTGTCTATGCCAAGAACAGCTGTTGTCTGAAGCATCAGTGGCGCGG 300
DB 82 ThrGlyLeuAlaLeuAlaTyrAlaIleGluGlnLeuPheAlaGlnAlaIleValaArg 101
QY 301 CCAGGGGTGCCCAAAGTCTGTGTGAGTGCAGATGAGCGGCTCCAGCACCTGTGGGC 360
DB 102 ProGlyValProLysValLeuValTyrValThrAspGlyGlySerSerAspProValGly 121
QY 361 CCCCCATGACAGAGACTCAAGGACCTGGCGGTCAACCGTTCATTTGATGACACCGGCGCA 420
DB 122 ProprometGlnGluLeuLeuAspLeuGlyValThrIlePheIleValSerThrGlyArg 141
QY 421 GGCACCTTCCTGGAGCTGTCAAGCCGCTGCTCAGCCCTGCGGAGAACACCTGCACTTT 480
DB 142 GlyAsnLeuLeuGluLeuLeuAlaAlaIleSerAlaProAlaGluLeuHisLeuHisPhe 161
QY 481 GTGACGTGATGATGCATCATTTGTCCAAAGCTGAGGGGCTCCATT 531
DB 162 ValaAspValaAspAspLeuProIleIleAlaArgGluLeuArgGlySerIle 178

RESULT 13
US-10-699-035A-31
Sequence 31, Application US/10699035A
Publication No. US20040214349A1
GENERAL INFORMATION:
APPLICANT: Bateman, John
TITLE OF INVENTION: A Molecular Marker
FILE REFERENCE: A36056 PCT USA A 071838.0142
CURRENT APPLICATION NUMBER: US/10/699,035A
PRIOR FILING DATE: 2003-10-31
PRIOR APPLICATION NUMBER: PCT/AU02/00542
PRIOR FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: AU PR4701/01
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 181
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: VA domain from WARP
US-10-699-035A-31
Alignment Scores:

```

Pred. No.: 1.83e-53 Length: 181
Score: 762.00 Matches: 149
Percent Similarity: 91.5% Conservative: 13
Best Local Similarity: 84.2% Mismatches: 15
Query Match: 72.6% Indels: 0
DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-31 (1-181)

```
Oy 1 GGGGACCTGATGTTCTCTGTCAGACGTCAGCCAGCAGCTCTCTCACTACGAGTTCTCCCG 60
   |||||.....|
Db 2 GlyAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 21
Oy 61 GTTCGGAGTTTGTGGGACAGCTGTGAGCTCCACTGCCCTGGGACCGGGGCGCTCGCT 120
   |||||.....|
Db 22 ValArgGluPheValIGlyGluLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 41
Oy 121 GCCAGCTGTGTGACAGTGGGACAGTGGCCATACACCGAGTTCCCTTGGGCGACAGCAGC 180
   |||||.....|
Db 42 AlaSerLeuValHisValIGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 61
Oy 181 TCGGCTGAGAGCTGCCAGAGATGGCGGTGCTCTGCCAGCGGATGGGTAGACCCAGC 240
   |||||.....|
Db 62 SerGlyGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 81
Oy 241 ACTGGCTGGCGCTGTCTATGTCACAGAACAGCTGTTTGTGAAACATCAGAGTCCCGG 300
   |||||.....|
Db 82 ThrGlyLeuAlaLeuAlaTyrAlaLysGluGluPheAlaGluGluAlaGlyAlaArg 101
Oy 301 CCAGGGGTGCCAAAGTGTGTGTGGGTGACAGATGGCGGCTCCAGGACCTCTGGGC 360
   |||||.....|
Db 102 ProGlyValProLysValLeuValIlePValThrAspGlyGlySerSerAspProValGly 121
Oy 361 CCCCCATGACAGAGCTCAAGAGCTGGGCGTCACCGGTGTTGATTTGTCAGCACCGCCGA 420
   |||||.....|
Db 122 ProPheMetGlnGluLeuLysAspLeuGlyValIleThrIlePheIleValSerThrGlyArg 141
Oy 421 GCGAAGCTTCCTGGAGCTGTACGCGCTGCTGAGCCCTGAGCCCGGAGAGACCTGCACATT 480
   |||||.....|
Db 142 GlyAsnLeuLeuGluLeuLeuAlaAlaAlaSerAlaProAlaGluLeuHisLeuHisPhe 161
Oy 481 GTGACCTGTGATGACCTGCACATCTATTGTCCAAAGCTGAGGGGCTCCATT 531
   |||||.....|
Db 162 ValAspValAspAspLeuProIleIleAlaArgGluLeuArgGlySerIle 178
```

RESULT 14

US-10-699-035A-4
; Sequence 4, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-4

Alignment Scores:
Pred. No.: 2.03e-53 Length: 415
Score: 762.00 Matches: 149
Percent Similarity: 91.5% Conservative: 13
Best Local Similarity: 84.2% Mismatches: 15

Query Match: 72.6% Indels: 0
DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-4 (1-415)

```
Oy 1 GGGGACCTGATGTTCTCTGTCAGACGTCAGCCAGCAGCTCTCTCACTACGAGTTCTCCCG 60
   |||||.....|
Db 33 GlyAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
Oy 61 GTTCGGAGTTTGTGGGACAGCTGTGAGCTCCACTGCCCTGGGACCGGGGCGCTCGCT 120
   |||||.....|
Db 53 ValArgGluPheValIGlyGluLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 72
Oy 121 GCCAGCTGTGTGACAGTGGGACAGTGGCCATACACCGAGTTCCCTTGGGCGACAGCAGC 180
   |||||.....|
Db 73 AlaSerLeuValHisValIGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 92
Oy 181 TCGGCTGAGAGCTGCCAGAGATGGCGGTGCTCTGCCAGCGGATGGGTAGACCCAGC 240
   |||||.....|
Db 93 SerGlyGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 112
Oy 241 ACTGGCTGGCGCTGTCTATGTCACAGAACAGCTGTTTGTGAAACATCAGAGTCCCGG 300
   |||||.....|
Db 113 ThrGlyLeuAlaLeuAlaTyrAlaLysGluGluPheAlaGluGluAlaGlyAlaArg 132
Oy 301 CCAGGGGTGCCAAAGTGTGTGTGGGTGACAGATGGCGGCTCCAGGACCTCTGGGC 360
   |||||.....|
Db 133 ProGlyValProLysValLeuValIlePValThrAspGlyGlySerSerAspProValGly 152
Oy 361 CCCCCATGACAGAGCTCAAGAGCTGGGCGTCACCGGTGTTGATTTGTCAGCACCGCCGA 420
   |||||.....|
Db 153 ProPheMetGlnGluLeuLysAspLeuGlyValIleThrIlePheIleValSerThrGlyArg 172
Oy 421 GCGAAGCTTCCTGGAGCTGTACGCGCTGCTGAGCCCTGAGCCCGGAGAGACCTGCACATT 480
   |||||.....|
Db 173 GlyAsnLeuLeuGluLeuLeuAlaAlaAlaSerAlaProAlaGluLeuHisLeuHisPhe 192
Oy 481 GTGACCTGTGATGACCTGCACATCTATTGTCCAAAGCTGAGGGGCTCCATT 531
   |||||.....|
Db 193 ValAspValAspAspLeuProIleIleAlaArgGluLeuArgGlySerIle 209
```

RESULT 15

US-10-699-035A-21
; Sequence 21, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-21

Alignment Scores:
Pred. No.: 2.03e-53 Length: 415
Score: 762.00 Matches: 149
Percent Similarity: 91.5% Conservative: 13
Best Local Similarity: 84.2% Mismatches: 15
Query Match: 72.6% Indels: 0
DB: 4 Gaps: 0

US-10-699-035A-1 (1-537) x US-10-699-035A-21 (1-415)

```
QY      1 GGGGACCTGATGTTCCCTGCTGACAGCTCAGCCAGCGTCTCTCACTACGAGTTCTCCCGG 60
Db      33 GlyAspLeuLeuPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArg 52
QY      61 GTTCGGGAGTTTGTGGGGCAGCTGGTGGCTCCACTGCCCTGGGACCCGGGGCCTGGCGT 120
Db      53 ValArgGluPheValGlyGluLeuValAlaThrMetSerPheGlyProGlyAlaLeuArg 72
QY      121 GCCAGCTGCTGTCACCTGGGACGTGGCCGATACCCGAGTTCCTCCCTGGCCGACACAGC 180
Db      73 AlaSerLeuValHisValGlySerGlnProHisThrGluPheThrPheAspGlnTyrSer 92
QY      181 TCGGGTGAAGCTGCCAGGATGCGGTGCTGCTTCTGCCCCAGCGCATGGGTGACACCCAC 240
Db      93 SerGlyGlnAlaIleArgAspAlaIleArgValAlaProGlnArgMetGlyAspThrAsn 112
QY      241 ACTGCTGCGCGCTGCTATGCTCAAGGAAACAGCTGTTTGTGAAGCATCAGGTGCCCGG 300
Db      113 ThrGlyLeuAlaLeuAlaTyrAlaLysGluGlnLeuPheAlaGluGluAlaArg 132
QY      301 CCAGGGGTGCCCAAGTGTGTTGGTGGGTGACAGATGGCGCTCCAGCGACCCGTGGGGC 360
Db      133 ProGlyValProLysValLeuValIlePValIleAspGlyGlySerSerAspProValGly 152
QY      361 CCCCCATGCAAGAGCTCAAGAGCTGGGCGTCAACCGTGTTCATTGTTCAGACCGGCCGA 420
Db      153 ProPheMetGlnGluLeuLysAspLeuGlyValThrIlePheIleValSerThrGlyArg 172
QY      421 GGCMACTTCTGAGAGCTGTCAAGCGGTGCTCAGCCCTGCGAGAGACACTGCACCTT 480
Db      173 GlyAsnLeuLeuGluLeuLeuAlaIleAlaSerAlaProAlaGluLysHisLeuHisPhe 192
QY      481 GTGAGCTGGATGACCTGCACATTCATTCACCAAGCTGAGGGGCTCCATT 531
Db      193 ValAspValAspAspLeuProIleIleAlaArgGluLeuArgGlySerIle 209
```

Search completed: February 13, 2006, 13:53:23
Job time : 91.701 secs

THIS PAGE BLANK (USPTO)


```
QY 226 ATGGGTGACACCCACACTGGCGCTGGCTGTATGCCAAAGAACAGCTGTTTCTGAA 285
Db 1 MetGlyAspThrHisThrGlyLeuAlaLeuValTyrAlaTyrGlnGlnPheAlaGln 20
QY 286 GCATCAGGTGCCCCGCGGAGGGGTGCCCAAGTGTGTGGTGTGGTACAGATGGCGGCTCC 345
Db 21 AlaSerGlyAlaArgProGlyValProLysValLeuValTyrValThrAspLysGlySer 40
QY 346 AGCGACCTGTGGGCGCCCCCATGACAGAGCTCAAGACCTGGGCGGTACCGCTTCAAT 405
Db 41 SerAspProValGlyProPrometGlnGlnLeuLysAspLeuGlyValThrValPheIle 60
QY 406 GTCAAGCACCGGCGGAGCACTTCTGTGAGCTGTGACCGCTGCTTCAAGCC 456
Db 61 ValSerThrGlyArg**AsnPheLeuGlnLeuSerAlaAlaAlaSerAla 77

RESULT 2
US-09-949-016-10340
; Sequence 10340, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10340
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10340

Alignment Scores:
Pred. No.: 6,256-15 Length: 584
Score: 259.00 Matches: 67
Percent Similarity: 51.1% Conservative: 25
Best Local Similarity: 37.2% Mismatches: 84
Query Match: 24.7% Indels: 4
DB: 2 Gaps: 2

US-10-699-035A-1 (1-537) x US-09-949-016-10340 (1-584)
QY 4 GACCTGATGTTCTGCTGAGACGCTCAGCCAGCGTCTCTCACTACGAGTTCTCCGGGTT 63
Db 348 AspLeuValLeuLeuValAspGlySerLysSerValAlaArgProGlnAsnPheGlnLeuVal 367
QY 64 CGGAGATTGTGGGCGACGCTGTGCTCCTCAGCTGCCCTGGGACCCGGGCGCTGCGTCC 123
Db 368 LysArgPheValAlaGlnIleValAspPheLeuAspValSerProGlnGlyThrArgVal 387
QY 124 AGTCTGTGACAGTGGGACGTGGCCATACACCGAGTTCCCTTGGCGGACAGACAGCTCG 183
Db 388 GlyLeuValGlnPheSerSerArgValArgThrGlnPheProLeuGlnYArgTyrGlyThr 407
QY 184 GGTGAGGCTGCCAGATGCGGTGCGTGTCTTCCGCCAGCGATGGGTGACACCCACACT 243
Db 408 AlaAlaGluValLysGlnAlaValLeuAlaValGluTyrMetGlnArgGlyThrMetThr 427
QY 244 GGCCTGGCGCTGTATGCAAGAACAGCTGTTTGTGAAGCATCAAGTCCCGGCA 303
Db 428 GlyLeuAlaLeuArgHisMetValGlnHisSerPheSerGlnAlaGlnGlyAlaArgPro 447
QY 304 -----GGGCTGCCCAAGAGTGTGTGTGGTGAAGATGGCGGCTCCAGCACCT 354
```

```
Db 448 ArgAlaLeuAsnValProArgValGlyLeuValPheThrAspGlyArgSerGlnAspAsp 467
QY 355 GTGGGCCCCCAGACAGACCTCAAGGACCTGGGCGGCGACGGTGTTCATGTCAGCAC 414
Db 468 IleSerValTyrAlaAlaArgAlaArgGlnGlnGlyIleValMetTyrAlaValGlyVal 487
QY 415 GCGCGAGGCAACTTCTGAGACCTGTACGCCCTGCTCAGCCCTGACCGCCGAGAACGACCTG 474
Db 488 GlyLysAlaValAlaGlnAlaGlnLeuArgGlnIleAlaSerGlnProAlaGlnLeuHisVal 507
QY 475 CACTTTGTG---GACGTGGATGACCTGCACATCATTTGTCCAAAGCTGAGGGGCTCCATT 531
Db 508 SerTyrAlaProAspPheGlyThrMetThrHisLeuGlnAsnLeuArgGlySerIle 527

RESULT 3
US-09-949-016-10341
; Sequence 10341, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10341
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10341

Alignment Scores:
Pred. No.: 6,256-15 Length: 584
Score: 259.00 Matches: 67
Percent Similarity: 51.1% Conservative: 25
Best Local Similarity: 37.2% Mismatches: 84
Query Match: 24.7% Indels: 4
DB: 2 Gaps: 2

US-10-699-035A-1 (1-537) x US-09-949-016-10341 (1-584)
QY 4 GACCTGATGTTCTGCTGAGACGCTCAGCCAGCGTCTCTCACTACGAGTTCTCCGGGTT 63
Db 348 AspLeuValLeuLeuValAspGlySerLysSerValAlaArgProGlnAsnPheGlnLeuVal 367
QY 64 CGGAGATTGTGGGCGACGCTGTGCTCCTCAGCTGCCCTGGGACCCGGGCGCTGCGTCC 123
Db 368 LysArgPheValAlaGlnIleValAspPheLeuAspValSerProGlnGlyThrArgVal 387
QY 124 AGTCTGTGACAGTGGGACGTGGCCATACACCGAGTTCCCTTGGCGGACAGACAGCTCG 183
Db 388 GlyLeuValGlnPheSerSerArgValArgThrGlnPheProLeuGlnYArgTyrGlyThr 407
QY 184 GGTGAGGCTGCCAGATGCGGTGCGTGTCTTCCGCCAGCGATGGGTGACACCCACACT 243
Db 408 AlaAlaGluValLysGlnAlaValLeuAlaValGluTyrMetGlnArgGlyThrMetThr 427
QY 244 GGCCTGGCGCTGTATGCAAGAACAGCTGTTTGTGAAGCATCAAGTCCCGGCA 303
Db 428 GlyLeuAlaLeuArgHisMetValGlnHisSerPheSerGlnAlaGlnGlyAlaArgPro 447
QY 304 -----GGGCTGCCCAAGAGTGTGTGTGGTGAAGATGGCGGCTCCAGCACCT 354
Db 448 ArgAlaLeuAsnValProArgValGlyLeuValPheThrAspGlyArgSerGlnAspAsp 467
```

QY 355 GTGGGCCCCCATGACAGAGCTCAAGACCTGGGGCTGACCGTGTCTATTGTACGACC 414
Db 468 lIeSerValtrPAlAlAlarGAlaYsGlUGlUGlYlIleValMeTyrAlaValGlYAl 487
QY 415 GCGCCAGGCAACTTCTGAGAGCTGACGCCGCTGCTTACGCCCTGCGGAGACACTG 474
Db 488 GlYlYsAlaValAlaGlAlaGlUleuArGluIleAlaSerGlUProAlaGlUleuHsVal 507
QY 475 CACTTTGTG---GAGCTGATGACCTGCACATCATGTGTCCAGAGAGTGAGGGCTTCATT 531
Db 508 SerTyrAlaProAspPheGlyThrMetThrHsleuLeuGluAsnLeuArGlySerIle 527

RESULT 4
US-09-919-497-57
Sequence 57, Application US/09919497
Patent No. 6773883
GENERAL INFORMATION:
APPLICANT: Muller, George L.
TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
FILE REFERENCE: B0801/7225
CURRENT APPLICATION NUMBER: US/09/919,497
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/221,735
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn version 3.0
SEQ ID NO 57
LENGTH: 755
TYPE: PRT
ORGANISM: Homo sapiens
US-09-919-497-57

Alignment Scores:
Pred. No.: 1,686-14 Length: 755
Score: 254.50 Matches: 53
Percent Similarity: 55.6% Conservative: 37
Best Local Similarity: 32.7% Mismatches: 71
Query Match: 24.3% Indels: 1
DB: 2 Gaps: 1

US-10-699-035A-1 (1-537) x US-09-919-497-57 (1-755)

QY 4 GACCTGATGTTCTCTGTCGACAGCTACGCCAGCTCTCACTACGAGTTCTCCGGATT 63
Db 7 AspleuValAlaPheMetValAspGlySerTrpSerIleGlyAspGluAsnPheAsnYsIle 26
QY 64 CGGAGATTGTGGGGCACTGTGGTCCACTG---CCCTGGGCAACCGGGCCCTGCGT 120
Db 27 lIeSerPheLeuTyrSerThrValGlYAlaLeuAsnYlIeGlYThrAspGlyThrGln 46
QY 121 GCCAGTCTGTCGACAGCTGGGCACTGCCCATACACCGATTCCTCCCTGCGCAGACAGC 180
Db 47 ValAlaMetValAlaGlnPheThrAspAspProArGlyThrGlnPheYsleuAsnAlaTyrYls 66
QY 181 TCGGGTGGAGGCTGCCAGGATGCGGTGCTGCTTGCACGAGCATGGGTGACACCCAC 240
Db 67 ThrYlsGlnThrLeuLeuAspAlaIleYsHsIleSerTyrIleYsGlyAsnThrYls 86
QY 241 ACTGGCTGGGCGCTGCTATGSCCAAGAAACAGCTGTTTGTGAAGCATCAGAGTCCGG 300
Db 87 ThrGlyYsAlaIleYsTyrValArGAspThrLeuPheThrAlaGlUleuSerGlyThrArG 106
QY 301 CCAGGGGTGCCCAAGATGCTGTGTGGGAGACAGAGGGGCTCCAGCAGACCTGGGGC 360
Db 107 ArGlyYlIeProYsAlaIleValAlaIleThrAspGlyArGserGlnAspAspValAsn 126
QY 361 CCCCCATGACAGAGCTCAAGACCTGGGCTCACCGCTTTCATTGTTCAGCACCGGCCGA 420
Db 127 lYsIleSerArGlnMetGlnleuAspGlyTyrSerIlePheAlaIleGlyValAlaAsp 146
QY 421 GGCMACTTCTGAGCTGTACGCCGCTGCTTACGCCCTGCGGAGAGACACTGCACCTTT 480

Db 147 AlaAspTyrSerGluLeuValSerIleGlySerYsProSerAlaArGHisValPhePhe 166
QY 481 GTGGAC 486
Db 167 ValAsp 168

RESULT 5
US-08-897-443-1
Sequence 1, Application US/08897443
Patent No. 5981263
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
APPLICANT: Kaser, Mathew
TITLE OF INVENTION: HUMAN MATRILIN-3
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,443
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0348 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: UTRSN02
CLONE: 681719
US-08-897-443-1

Alignment Scores:
Pred. No.: 8,396-14 Length: 638
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
DB: 1 Gaps: 2

US-10-699-035A-1 (1-537) x US-08-897-443-1 (1-638)

QY 4 GACCTGATGTTCTCTGTCGACAGCTACGCCAGCTCTCTCACTACGAGTTCTCCGGATT 63
Db 57 AspleuValAlaPheIleAlaAspSerIleArGserValAsnThrHsAspTyrAlaYsVal 76
QY 64 CGGAGATTGTGGGGCACTGTGGTCCACTGACCGCTGGGCAACCGGGCCCTGCGTGC 123
Db 77 lYsGlnPheIleValAlaSpIleleuGlnPheLeuAspIleGlyProAspValThrArGVal 96
QY 124 AGTCTGTGACAGTGGGAGTGGGCAATACACCGAGTTCCCTTGGGCCAGACAGCTGC 183

```
Db 97 GlyLeuLeuGlnTyrGlySerThrValLysAsnGlnPheSerLeuLysThrPheLysArg 116
Oy 184 GGTGAGCGTGCAGGATGCGGTGCGCTTCTGCGCCAGCGATGGTGACACCCACT 243
Db 117 LysSerGluValGluValArgAlaValLysArgMetArgHisLeuSerThrGlyThrMetThr 136
Oy 244 GGCCTGGCGCTGCTATGCAAGACAGCTGTTTGCTGAAGCATCAGTGCCCGGCA 303
Db 137 GlyLeuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluAlaGluGlyAlaArgPro 156
Oy 304 -----GGGTCGCCCAAGTGTGTGTGCGGTGCGGACAGATGGCGGCTCCAGCACCT 354
Db 157 LeuArgGluAsnValProArgValIleMetIleValThrArgGlyArgProGlnAspSer 176
Oy 355 GTGGGCCCCCAGTCCAGACAGCTGAGACCTGGCGGCGTCACTGTTGATTCAGCACCC 414
Db 177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
Oy 415 GCGCGAGGCAACTTCCTGAGAGCTGTCAAGCCGCTGCTCAGCCCTGCGGAGAGCACCTG 474
Db 197 GlyGlnValAspPheAsnThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216
Oy 475 -----CACTTTGTGACCTGATGACCTTCGACATCATTTGTCGAAGAG 516
Db 217 PheLeuValAlaAsnPheserGlnIleGlnThrLeuThrSerValPheGlnLys 234

RESULT 6
US-09-907-794A-34
Sequence 34, Application US/09907794A
Patent No. 6635468
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Steward, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,794A
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
```

```

; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-794A-34

Alignment Scores:
Pred. No.: 9.18e-14 Length: 915
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Gaps: 2
DB: 2

US-10-699-035A-1 (1-537) x US-09-907-794A-34 (1-915)
Oy 4 GACCTGATGTCCTGCTGAGACGCTGACGCGTCTCACTAGAGTCTCCGGGTT 63
Db 57 AspLeuValPheIleIleAspSerSerArgSerValAsnThrHisAspTyrAlaLysVal 76
Oy 64 CGGAGTTTGTGGGAGACCTGTGCTCCACTGCCCCCTGGACCGGGGCGCTGCTGCC 123
Db 77 LysGluPheIleValAspIleLeuGlnPheLeuAspIleGlyProAspValThrArgVal 96
Oy 124 AGCTGTGTCACGTGGGACGTGCGGCATACCCGAGTTCCCTTGGCCAGACAGCTCG 183
Db 97 GlyLeuLeuGlnTyrGlySerThrValLysAsnGlnPheSerLeuLysThrPheLysArg 116
Oy 184 GGTGAGCGTGCAGGATGCGGTGCGCTTCTGCGCCAGCGATGGTGACACCCACT 243
Db 117 LysSerGluValGluValArgAlaValLysArgMetArgHisLeuSerThrGlyThrMetThr 136
Oy 244 GGCCTGGCGCTGCTATGCAAGACAGCTGTTTGCTGAAGCATCAGTGCCCGGCA 303
Db 137 GlyLeuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluAlaGluGlyAlaArgPro 156
Oy 304 -----GGGTCGCCCAAGTGTGTGTGCGGTGCGGACAGATGGCGGCTCCAGCACCT 354
Db 157 LeuArgGluAsnValProArgValIleMetIleValThrArgGlyArgProGlnAspSer 176
Oy 355 GTGGGCCCCCAGTCCAGACAGCTGAGACCTGGCGGCGTCACTGTTGATTCAGCACCC 414
Db 177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
Oy 415 GCGCGAGGCAACTTCCTGAGAGCTGTCAAGCCGCTGCTCAGCCCTGCGGAGAGCACCTG 474
Db 197 GlyGlnValAspPheAsnThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216
Oy 475 -----CACTTTGTGACCTGATGACCTTCGACATCATTTGTCGAAGAG 516
```

Db 217 PheLeuValAlaAsnPheSerGlnIleGluThrLeuThrSerValPheGlnIlys 234

RESULT 7
US-09-905-125A-34
Sequence 34, Application US/09905125A
Patent No. 6664376
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Bacon, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905,125A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423

```

; SEQ ID NO 34
;
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-125A-34

```

Alignment Scores:	
Pred. No.:	9.18e-14
Score:	246.50
Percent Similarity:	52.2%
Best Local Similarity:	32.0%
Query Match:	23.5%
DB:	2
Length:	915
Matches:	57
Conservative:	36
Mismatches:	78
Indels:	7
Gaps:	2

US-10-699-035A-1 (1-537) X US-09-905-125A-34 (1-915)

```

OY      GACCTATGTCCTCGTGGAGACGCTCAGCCAGCGCTCTCATTACGAGTTCTCCGGGTT 63
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      AsplneValaPheIleIleAspserIerIgservalAsnThrIshApTyfAlaVal 76
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
OY      CGGAGATTGTGGGGCAGCTGTGCTCCTCACTGCCCCGTGGGACACGGGGCCCTCGGCC 123
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      77  LysGIuPheIleValAspIleIeuuInPheIleuAspIleGIyProAspValThrArgVal 96
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
OY      124 AGTCTGTGACGTGGGCGAGTGTGGCCATACACCGAGTTCCCTCTGGCCAGCACAGCTCG 183
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      97  G1LeuIleuGlnItyrGIySerThrValIyAsnGIuPheSerIleuTythrPheIyAsr 116
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
OY      184 GCGAGGCGCCCGCAGGATGCGGTGGGTCTTCGCTCCGACCGCATATGGGTGACACCCACACT 243
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      117 LysSerGIuValGIuIrgAlaValIyAsrIgmethArghIsIeuserIthrngIyThrMetThr 133
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
OY      244 GGCCTGCGCGCTGCTGTATGCCAAGACAGCTGTTTGCTGAAGCATCAGGTGCCGCGCA 303
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      137 G1LeuAlaIleGIuItyrAlaIeuAsnIleAlaPheSerGIuAlaGIuGIyAlaArgPro 154
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
OY      304 -----GGGGTGGCCAAAGTGTGCGGTGGGTGGGAGACATGGCGCGCTCCAGGACGCT 354
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      157 LeuArgGIuAsnAlaProArgValIleMetIleValInrAspGIyArgProGlnAspSer 176
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
OY      355 GTGGGCCCCCCTATGCAGAGCTCAAGACCTGGGCGGTCAACGTTTCATTGTGCAGACCC 414
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      177 ValAlaGIuValAlaIaIaIyAlaArgAspThrngIyIleuIlePheAlaIleGIyVal 196
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
OY      415 GCGCGAGGCAACTTCTGTGAGCTGTTCAGCCGCTGCTTACGCCCTCGCCGGAAGACACTG 478
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      197 G1GIuIValAspPheAsnThrIleuIyAsnIerIleGIySerGIuProHISGIuAspHisVal 216
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
OY      475 -----CACTTTGTGACGTGGATGCATGTCGCATCATTTGTTCACAGAG 516
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      217 PheIuValAlaAsnPheSerGIuIleGIuIthrIleuThrSerValPheGlnIyS 234
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 8
US-09-902-775A-34
; Sequence 34, Application US/09902775A
; Patent No. 6686451
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerltisen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austen L.
; APPLICANT: Hillan, Kenneth, J.

```

```

: APPLICANT: KlJavin, Ivar J.
: APPLICANT: Mather, Jennie P.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: 10466-14
: CURRENT APPLICATION NUMBER: US/09/902,775A
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: PCT/US99/20944
: PRIOR FILING DATE: 1999-09-13
: PRIOR APPLICATION NUMBER: PCT/US99/21090
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/21547
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/23089
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: PCT/US99/28214
: PRIOR FILING DATE: 1999-11-29
: PRIOR APPLICATION NUMBER: PCT/US99/28313
: PRIOR FILING DATE: 1999-11-30
: PRIOR APPLICATION NUMBER: PCT/US99/28564
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/28565
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30095
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: PCT/US99/30911
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US99/30999
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 423
: SEQ ID NO 34
: LENGTH: 915
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-902-775A-34

Alignment Scores:
Pred. No.: 9,18e-14 Length: 515
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
DB: 2 Gaps: 2

US-10-699-035A-1 (1-537) x US-09-902-775A-34 (1-915)

QY 4 GACCTGATGTTCTGCTGACAGCTTACGCCAGCTTCTCACTAGCAGTTTCCCGGTT 63
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 57 AspluValAlpheilleAspserSerValAsnThHisAspTylAlaVal 76
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 64 CGGAGTTGTGGGAGCGAGCTGCAGTCCCGCCCGGAGCCGCGGCGCTGCGTCC 123
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 77 LysGluPheIleValAspIleuGlnPheLeuAspIleGlyProAspValThrArgVal 96
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```

QY 124 AGCTGTGCACGTGGGCGAGTCGGCCATACCCGAGTTCCCTTGGCCAGACAGCTCG 183
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 GlyLeuIeuGlnTyrGlySerThrValLysAsnGluPheSerLeuYrThrPheLysArg 116
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 184 GGTAGAGCTGCCCGAGATGGCGGTGCGTCTTCTGCGCCAGCGGATGGGAGACCCACACT 243
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117 LysSerGluValAlGluArgAlaValLysArgMetArgHisLeuSerThrGlyThrMetThr 136
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 244 GGCCTGGCGCTGTCTATATGCCAAGAACAGCTGTTTGGTGAAGATCAGTGCCCGGCA 303
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 137 GlyLeuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluAlaGlnGlyAlaArgPro 156
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 304 -----GGGTTGCCAAGTGTCTGTGGTGTGACAGATGGCGGCTCCAGCAGCCCT 354
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 LeuArgGluAsnValProArgValIleMetIleValThrArgGlyArgProGlnAspSer 176
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 355 GTGGGCCCCCAGTGCAGAGACTCAAGACCTGGGCGGCACCGTTCATTGTCAAGACC 414
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 415 GCGCGAGCACTTCTGAGCTGTGCAGCCCGCTGCTCAGCCCTGCGCGAGACAGCCTCG 474
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197 GlyIuValAspPheAsnThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 475 -----CACTTTGTGACGTGATGATGATCCTGCACATCTTGTCCAAAG 516
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 217 PheLeuValAlaAsnPheSerGlnIleGluThrLeuThrSerValPheGlnLys 234
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-906-700-34
: Sequence 34, Application US/09906700
: Patent No. 6723535
: GENERAL INFORMATION:
: APPLICANT: Genentech, Inc.
: APPLICANT: Ashkenazi, Avi
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Baton, Dan L.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Geider, Hanspeter
: APPLICANT: Gerltsen, Mary B.
: APPLICANT: Goddard, A.
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth, J.
: APPLICANT: KlJavin, Ivar J.
: APPLICANT: Mather, Jennie P.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: 10466-14
: CURRENT APPLICATION NUMBER: US/09/906,700
: CURRENT FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
```

```

/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ PRIOR APPLICATION NUMBER: PCT/US99/28313
/ PRIOR FILING DATE: 1999-11-30
/ PRIOR APPLICATION NUMBER: PCT/US99/28564
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/28565
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/30095
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: PCT/US99/30911
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US99/30999
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US00/00219
/ PRIOR FILING DATE: 2000-01-05
/ NUMBER OF SEQ ID NOS: 423
/ SEQ ID NO 34
/ LENGTH: 915
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-906-700-34
```

```

Alignment Scores:
Pred. No.: 9,18e-14 Length: 515
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
DB: 2 Gaps: 2
```

US-10-699-035A-1 (1-537) x US-09-906-700-34 (1-915)

```

QY 4 GACCTGATTTCTCTGTGACAGCTGACGAGCTTCTACAGCTTCTCCGGGTT 63
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 57 ApleuValApleuValApleuValApleuValApleuValApleuValApleuVal 76
QY 64 CGGAGTTTGTGGGAGCTGTGGCTCAGTCCCTCGGAGCCGGGCGCTGGCC 123
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 77 LysGluPheIleValApleuValApleuValApleuValApleuValApleuVal 96
QY 124 AGTCTGATGACAGTGTGGGAGCTGACGAGCTTCTACAGCTTCTCCGGGTT 183
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 97 GlyLeuLeuGlnTyrClySerThrValLysGlnGlnPheSerLeuLysThrPheLysVal 116
QY 184 GGTGAGGCTCCAGAGATCGGTGCTGCTTGTCCGAGCGATGGGTACACCACT 243
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 117 LysSerGluValApleuValApleuValApleuValApleuValApleuVal 136
QY 244 GGGCTGGGCTGTGTATGCAAGAGAACGCTGTTTGCAGACATCAGTCCGGGCA 303
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 137 GlyLeuAlaIleGlnTyrAlaLeuValApleuValApleuValApleuValApleuVal 156
QY 304 -----GGGCTGGGCTGTGTATGCAAGAGAACGCTGTTTGCAGACATCAGTCCGGGCA 354
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 157 LeuArgGluValApleuValApleuValApleuValApleuValApleuVal 176
QY 355 GTGGGCCCCCATGACGAGCTCAAGAGACCTGGGCGGTACCGTGTTCATTGTCAAGCACCC 414
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 177 ValAlaGluValAlaAlaLysValApleuValApleuValApleuValApleuVal 196
QY 415 GGGCAGGAGAACTTCTGTGAGCGGTGACGGCGCTTACCGCTGCGAGAACCGG 474
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 197 GlyGlnValApleuValApleuValApleuValApleuValApleuValApleuVal 216
```

```

QY 475 -----CACTTGTGACGTGATGACTGACATCATTTCTCAAGAG 516
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 217 PheLeuValAlaApleuValApleuValApleuValApleuValApleuVal 234
RESULT 10
US-09-903-603A-34
/ Sequence 34, Application US/09903603A
/ Patent No. 6767995
/ GENERAL INFORMATION:
/ APPLICANT: Genentech, Inc.
/ APPLICANT: Ashkenazi, Avi
/ APPLICANT: Botstein, David
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, A.
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.
/ APPLICANT: Gueney, Austin L.
/ APPLICANT: Hillan, Kenneth, J.
/ APPLICANT: Kijavlin, Ivar J.
/ APPLICANT: Mather, Jennie P.
/ APPLICANT: Pan, James
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William, I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ FILE REFERENCE: GNE.1618P2C12
/ CURRENT APPLICATION NUMBER: US/09/903,603A
/ PRIOR FILING DATE: 2001-07-11
/ PRIOR APPLICATION NUMBER: PCT/US00/04414
/ PRIOR FILING DATE: 2000-02-22
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
/ PRIOR APPLICATION NUMBER: US 60/146,222
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: PCT/US99/20594
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ PRIOR APPLICATION NUMBER: PCT/US99/28313
/ PRIOR FILING DATE: 1999-11-30
/ PRIOR APPLICATION NUMBER: PCT/US99/28564
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/28565
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/30095
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: PCT/US99/30911
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US99/30999
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US00/00219
/ PRIOR FILING DATE: 2000-01-05
```

```

; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-603A-34

Alignment Scores:
Pred. No.: 9,18e-14      Length: 915
Score: 246.50           Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5%      Indels: 7
DB: 2                   Gaps: 2

US-10-699-035A-1 (1-537) x US-09-903-603A-34 (1-915)
QY 4 GACCTGATGTCCTGCTGACAGCTCAGCCAGCTCTCTCACTAGAGTTCTCCGGGTT 63
Db 57 AspleuValPheilelleAspserSerArgSerValAsnThrHisAspTyrAlaVal 76
QY 64 CGGAGATTGTGGGGACAGCTGAGCTCCACTGCTGGGACCGGGGCTGGTCC 123
Db 77 LysGluPheilelleValAspIleuGlnPheleuAspIleGlyProAspValThrArgVal 96
QY 124 AGCTGTGTCAGCGTGGGACAGCTGCGCCATACACCGAGTTCCCTTGGCCAGACAGCTCG 183
Db 97 GlyleuLeuGlnTyrGlySerThrValLysAsnGluPheSerLeuLysThrPheLysArg 116
QY 184 GGTAGGCTGCCAGAGATGCGGTGCTGTTGTGCCACGAGATGGGTGACCCCACT 243
Db 117 LysSerGluValAlaArgAlaValAlaLysArgMetArgHisLeuSerThrGlyThrMetThr 136
QY 244 GACCTGAGCTGTGCTATGCTGACAGGACAGCTGTTGCTGTAAGCATCAGTCCCGGCA 303
Db 137 GlyleuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluAlaGluLysAlaArgPro 156
QY 304 -----GGGGTCCCAAGTGTGCTGTTGTGCTGACAGATGGGCGCTCCAGCGACT 354
Db 157 LeuArgGluAsnValProArgValIleMetIleValThrAspGlyArgProGlnAspSer 176
QY 355 GTGGGCCCCCATGACAGAGCTCAAGACCTGGGCGCTCAGCGTTCATTGTTCAGCAC 414
Db 177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyLeuIlePheAlaIleGlyVal 196
QY 415 GCGCCAGGCAACTCTCTGAGCTGTGAGCGCTGCTCAGCCCTGCGGAGAGACCTG 474
Db 197 GlyGluValAspPheAsnThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216
QY 475 -----CACTTGTGAGCGTGATGACCTGACATCATTTGTCACAGAG 516
Db 217 PheLeuValAlaAsnPheSerGlnIleGluThrLeuThrSerValPheGlnLys 234

RESULT 11
US-09-904-920A-34
; Sequence 34, Application US/09904920A
; Patent No. 6806352
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
```

```

; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Macher, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas P.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,920A
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-904-920A-34

Alignment Scores:
Pred. No.: 9,18e-14      Length: 915
Score: 246.50           Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5%      Indels: 7
DB: 2                   Gaps: 2

US-10-699-035A-1 (1-537) x US-09-904-920A-34 (1-915)
QY 4 GACCTGATGTCCTGCTGACAGCTCAGCCAGCTCTCTCACTAGAGTTCTCCGGGTT 63
Db 57 AspleuValPheilelleAspserSerArgSerValAsnThrHisAspTyrAlaVal 76
QY 64 CGGAGATTGTGGGGACAGCTGAGCTCCACTGCTGGGACCGGGGCTGGTCC 123
Db 77 LysGluPheilelleValAspIleuGlnPheleuAspIleGlyProAspValThrArgVal 96
```

```

QY      124 AGTGTGACAGCTGCGAGTCGCGCCATACACCGAGTTCCTTCGCGCAGCAGCAGCTCG 183
      |||:::|||||
Db      97 GlyLeuLeuGlnTyrGlySerThrValLysAsnGlnPheSerLeuYsrThrPheLysArg 116
QY      184 GGTGAGCTGCCCGAGATGCGGTGCTGCTTGTGCCAGGAGTGGGTGACACCCACT 243
      |||:::|||||
Db      117 LysSerGluValGlnLysArgMetArgHisLeuSerThrGlyThrMetThr 136
QY      244 GGCCTGCGCTGCTATGCGCAAGAACAGCTTTGTCGAACATCAGTGGCCGCGCA 303
      |||:::|||||
Db      137 GlyLeuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluValGlnGlyAlaArgPro 156
QY      304 -----GGGCTGCCAAGTGTGCTGCGGTGACAGATGGCGCTCCAGCAGCCT 354
      |||:::|||||
Db      157 LeuArgGluAsnValProArgValIleMetIleValThrAspGlyArgProGlnAspSer 176
QY      355 GTGGGCCCCCATGACAGAGCTCAAGACCTGGCGGTACCGTTCATTGTACAGCACC 414
      |||:::|||||
Db      177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
QY      415 GGCAGGAGCACTTCTGTGAGCTGTGACGCGCTGCGCCCTGCGCAGAGACCTG 474
      |||:::|||||
Db      197 GlyGlnValAspPheAsnThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216
QY      475 -----CACTTGTGACGTGTGATGATGACCTGACATCATTTGTCCAGAG 516
      |||:::|||||
Db      217 PheLeuValAlaAsnPheSerGlnIleGluThrLeuThrSerValPheGlnHis 234

RESULT 12
US-09-909-064-34
; Sequence 34, Application US/09909064
; Patent No. 6818449
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvarolf, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Guiney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,064
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594

```

```

; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-064-34

Alignment Scores:
Pred. No.: 9,186-14 Length: 915
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatch: 78
Query Match: 23.5% Indels: 7
DB: 2 Gaps: 2

US-10-699-035A-1 (1-537) x US-09-909-064-34 (1-915)
QY      4 GACCTGATGTCCTGCTGACAGCTCAGCAGCTCTCTCACTACAGATTCTCCGGATT 63
      |||:::|||||
Db      57 AspLeuValPheIleIleAspSerSerArgSerValAsnThrHisAspTyrAlaLysVal 76
QY      64 CGGAGTTTGTGGGAGCTGTGCTCCACTGCCCCCTGGGACCGGGCCCTGCTGCC 123
      |||:::|||||
Db      77 LysGluPheIleValAlaLysIleLeuGlnPheLeuAspIleGlyProAspValThrArgVal 96
QY      124 AGTGTGACAGCTGCGAGTCGCGCCATACACCGAGTTCCTTCGCGCAGCAGCTCG 183
      |||:::|||||
Db      97 GlyLeuLeuGlnTyrGlySerThrValLysAsnGlnPheSerLeuYsrThrPheLysArg 116
QY      184 GGTGAGCTGCCCGAGATGCGGTGCTGCTTGTGCCAGGAGTGGGTGACACCCACT 243
      |||:::|||||
Db      117 LysSerGluValGlnLysArgMetArgHisLeuSerThrGlyThrMetThr 136
QY      244 GGCCTGCGCTGCTATGCGCAAGAACAGCTTTGTCGAACATCAGTGGCCGCGCA 303
      |||:::|||||
Db      137 GlyLeuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluValGlnGlyAlaArgPro 156
QY      304 -----GGGCTGCCAAGTGTGCTGCGGTGACAGATGGCGCTCCAGCAGCCT 354
      |||:::|||||
Db      157 LeuArgGluAsnValProArgValIleMetIleValThrAspGlyArgProGlnAspSer 176
QY      355 GTGGGCCCCCATGACAGAGCTCAAGACCTGGCGGTACCGTTCATTGTACAGCACC 414
      |||:::|||||
Db      177 ValAlaGluValAlaAlaLysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196
QY      415 GGCAGGAGCACTTCTGTGAGCTGTGACGCGCTGCGCCCTGCGCAGAGACCTG 474
      |||:::|||||
Db      197 GlyGlnValAspPheAsnThrLeuLysSerIleGlySerGluProHisGluAspHisVal 216

```

Qy 475 -----CACTTGTGACGTGATGACCTGCACATCATTTGCCAAG 516
Db 217 PheleuValAlaAsnPheserGlnIleGluThrLeuThrSerValPheGlnIlys 234

RESULT 13
US-09-905-381A-34
; Sequence 34, Application US/09905381A
; Patent No. 6818746
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Guiney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,381A
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219

; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-381A-34

Alignment Scores:
Pred. No.: 9,18e-14 Length: 915
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
DB: 2 Gaps: 2

US-10-699-035A-1 (1-537) x US-09-905-381A-34 (1-915)

Qy 4 GACCTGATGTTCTGCTGACGACCTGACCGCTCTCTACACTACGAGTTCTCCGGGTT 63
Db 57 AspleuValPheIleIleAspserserArgserValaThrH1sAspTyrAlaIysVal 76

Qy 64 CGGAGTTTGGGGGACGCTGATGCTCCACTGCGCCCTGGGACCGGGGCGCTGGCTCC 123
Db 77 LysGluPheIleValAspIleuGlnPheLeuAspIleGlyProAspValThrArgVal 96

Qy 124 AGCTGTGACAGTGGGAGCTCGGCCATACACGAGTTCCTCCCTCGCCACAGCTCG 183
Db 97 GlyLeuLeuGlnTyrGlyserThrValYsAsnGluPheSerLeuYThrPheYsArg 116

Qy 184 GGTGAGCTGCCAGAGATCGCGTGCCTTCTGCCCCAGGCGATGGGACACCCACT 243
Db 117 LysSerGluValGluArgAlaValIysArgMetArgH1sLeuSerThrGlyThrMetThr 136

Qy 244 GGCCCTGGCGCTGCTGCTAGAGACAGGCTGTTGTGAAGATCAGTGGCCGGGCA 303
Db 137 GlyLeuAlaIleGlnTyrAlaLeuAsnIleAlaPheSerGluAlaGluGlyAlaArgPro 156

Qy 304 -----GGGATGCCCAAGTGTGTGTGGTGAAGATGGGCGCTCCAGCACCT 354
Db 157 LeuArgGluAsnValProArgValIleMetIleValThrAspGlyArgProGlnAspSer 176

Qy 355 GTGGGCCCCCAGACGACGACTCAAGACCTGGGCGCTACCGCTTATTTGTACAGCAC 414
Db 177 ValAlaGluValAlaAlaIysAlaArgAspThrGlyIleLeuIlePheAlaIleGlyVal 196

Qy 415 GGCGGAGGACACTTCTGAGCTGTGACCGCTGCTGAGCCCGCTGGCCGAGAGACCTG 474
Db 197 GlyIleValAspPheAsnThrLeuIysSerIleGlySerGluProH1sGluAspHisVal 216

Qy 475 -----CACTTGTGACGTGATGACCTGCACATCATTTGCCAAG 516
Db 217 PheleuValAlaAsnPheserGlnIleGluThrLeuThrSerValPheGlnIlys 234

RESULT 14
US-09-906-618-34
; Sequence 34, Application US/09906618
; Patent No. 6828146
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.

PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-906-646-34

Alignment Scores:

Pred. No.: 9,18e-14 Length: 915
Score: 246.50 Matches: 57
Percent Similarity: 52.2% Conservative: 36
Best Local Similarity: 32.0% Mismatches: 78
Query Match: 23.5% Indels: 7
DB: 2 Gaps: 2

US-10-699-035A-1 (1-537) x US-09-906-646-34 (1-915)

QY 4 GACCTGATGTTCTCTGACAGCTGACAGCTGTTCTCACTAGAGTTCTCCGGGTT 63
DB 57 AspleuValPheIleIleAspSerSerIArgSerValAsnThrHisAspTyrAlaIysVal 76
QY 64 CGGAGTTTGTGGGACAGCTGGTCCACTGCCCCCTGGGACCCGGGACCTGGCTGCC 123
DB 77 LysGIuPheIleValAspIleuGlnPheIuAspIleGIyProAspValThrArgVal 96
QY 124 AGTCGTGACAGCTGGGACAGTGGCCATACACGAGTTCCTCCCTGGCCAGACAGCTCG 183
DB 97 GlyLeuLeuGlnTyrGlySerThrValIysAsnGIuPheSerLeuIlyThrPheIysArg 116
QY 184 GGTGAGGCTGCCAGATGCGGTGCGCTGCTTGTGCCAGCGATGATGACACCCACT 243
DB 117 LysSerGIuValAlaIysArgAlaValIysArgMetArgHisIleuSerThrGIyThrMetThr 136
QY 244 GGCCTGGCGCTGTCTATGCCAAGAAACAGCTGTTTCTGAAGCATCAGGTGCCGCCA 303
DB 137 GlyLeuAlaIleGIuTyrAlaLeuAsnIleAlaPheSerGIuAlaGIuGIyAlaArgPro 156
QY 304 -----GGGCTGCCAAAGTGCCTGCTGCTGATGACAGATGGCGGCTCCAGCACCT 354
DB 157 LeuArgGIuAsnValProArgValIleMetIleValIThrAspGIyArgProGlnAspSer 176
QY 355 GTGGACCCCCCATGACAGAGCTCAAGAGCATGGAGCGTCAACCGTTCATTGTACAGACC 414
DB 177 ValAlaGIuValAlaAlaIysAlaIysArgThrGIyIleuIlePheAlaIleGIyVal 196
QY 415 GCGCAGGCACTTCTGGAGCTGTCAAGCCGCTGCCTCAGCCCTCGCAGAAAGACCTCG 474
DB 474

DB 197 GlyGIuValAspPheAsnThrLeuIysSerIleGIySerGIuProHisGIuAspHisVal 216
QY 475 -----CACTTGTGACGTGATGACCTGCACATCATTTGTCCAAAG 516
DB 217 PheLeuValAlaAsnPheSerGlnIleGIuThrIleuThrSerValPheGlnIys 234

Search completed: February 13, 2006, 13:43:34
Job time : 25.7889 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:23:58 ; Search time 10.0124 Seconds
(without alignments)
2410.129 Million cell updates/sec

Title: US-10-699-035A-5

Perfect score: 2380
Sequence: 1 atgtccctccgtgagcgagcgcct.....ccgcacgcgtgagcgctaa 1254

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p model -DEV=xldp
-Q=/abs/ABSSWEB.spool/US10699035/runat.13022006.062444.25463/app.query.fasta_1
-DB=PIR -QFMT=fastan -SUFPTX=tpir -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=q
-UNITS=bites -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -DIST=45
-DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs802p
-USER=US10699035 @CCN 1.1.77 @runat.13022006.062444.25463 -NCPU=6 -ICU=3
-NO_MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: PIR 80:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	444	18.7	3124	2 A40020	collagen alpha 1(X
2	387	16.3	2944	2 A54849	collagen alpha 1(V
3	382	16.1	1857	2 S31212	collagen alpha 1(X
4	382	16.1	1888	2 S78476	collagen alpha 1(X
5	371	15.6	1747	2 A45974	collagen alpha 1(X
6	272.5	11.4	929	2 F51027	type XII collagen
7	268	11.3	843	2 A40970	undulin 1 - human
8	263	11.1	660	1 Q08B3	BHLFI protein - hu
9	248.5	10.4	493	2 A33809	cartilage matrix p
10	246.5	10.4	1106	2 J00405	hypothetical 119.5
11	243.5	10.2	500	2 S66522	cartilage matrix p
12	243.5	10.2	660	1 Q08B3	BHLFI protein - hu
13	240	10.1	1344	1 A35175	mucin 1 precursor,
14	233.5	9.8	496	2 A37979	cartilage matrix p

15	230.5	9.7	3137	2 A37797	collagen alpha 3(V
16	227	9.5	3176	2 CGH03A	collagen alpha 3(V
17	226	9.5	724	2 A48569	antigen Em100 - E1
18	221.5	9.3	460	2 T33110	hypothetical prote
19	221	9.3	1446	1 A45344	immediate-early pr
20	219	9.2	1042	1 CGCH1S	collagen alpha 1(I)
21	218.5	9.2	1460	1 EDBE1F	immediate-early pr
22	218	9.2	1042	1 CGCH1S	collagen alpha 1(I)
23	217.5	9.1	779	1 CGB01S	collagen alpha 1(I)
24	217	9.1	1027	2 S28774	collagen alpha cha
25	216.5	9.1	1453	2 S21626	collagen alpha 1(I)
26	216	9.1	1151	2 T18535	high molecular mas
27	216	9.1	1173	1 A43291	collagen alpha 2(I)
28	216	9.1	1497	2 I49607	procollagen type V
29	215	9.0	1496	1 CGH02V	collagen alpha 2(V
30	214	9.0	964	1 CGCH2S	collagen alpha 2(I)
31	213.5	9.0	712	2 A45638	immunodominant mic
32	213	8.9	741	2 T46488	hypothetical prote
33	212.5	8.9	1151	2 A45226	integrin alpha-1 c
34	212.5	8.9	1366	1 CGH02S	collagen alpha 2(I)
35	212	8.9	886	2 I50694	collagen alpha 1(I)
36	212	8.9	1464	1 CGH01S	collagen alpha 1(I)
37	211	8.9	1464	1 CGH01S	collagen alpha 1(I)
38	210.5	8.8	1466	1 CGH07L	collagen alpha 1(I)
39	210	8.8	1453	2 S21626	collagen alpha 1(I)
40	209.5	8.8	1414	1 S23809	collagen alpha 2(I)
41	209	8.8	1414	1 S23809	collagen alpha 2(I)
42	208.5	8.8	779	1 CGB01S	collagen alpha 1(I)
43	208.5	8.8	1460	1 EDBE1F	immediate-early pr
44	207.5	8.7	1446	1 A45344	immediate-early pr
45	207.5	8.7	1958	2 B40505	hypothetical prote

ALIGNMENTS

RESULT 1
A40020
collagen alpha 1(XII) chain precursor - chicken
N:Alternate names: fibrochimerin
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #ext change 09-Jul-2004
C:Accession: A40020; A54485; B34485; A28037; S23814; S22554; S28811
J:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Oba
R.; Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
U. Biol. Chem. 264, 19772-19778, 1988
A:Title: Type XII collagen. A large multidomain molecule with partial homology to type
A:Reference number: A34485; MUID:90062079; PMID:2584192
A:Accession: A34485
A:Molecule type: mRNA
A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
A:Cross-references: UNIPARC:UPI0000171233; EMBL:U05137; NID:G211284; PIDN:AAA48635.1; P
A:Accession: B34485
A:Molecule type: protein
A:Residues: 2772-2792; 2846-2873 <GOR2>
A:Cross-references: UNIPARC:UPI0000173CF; UNIPARC:UPI0000173C40
R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A:Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA
A:Reference number: A28037; MUID:87317590; PMID:3476925
A:Accession: A28037
A:Molecule type: mRNA
A:Residues: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>
A:Cross-references: UNIPARC:UPI00001712P4; EMBL:M17375; NID:G211649; PIDN:AAA48718.1; P
A:Note: this sequence has been revised in reference A34485

R:Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992
A:Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A:Reference number: S23814; MUID:92362621; PMID:1323460
A:Accession: S23814
A:Molecule type: protein
A:Residues: 'X',1333,'Q',1335-1347,1914-1928,2504,'X',2506,'X',2508-2511,'X',2513-2517
A:Cross-references: UNIPARC:UPI0000173C41; UNIPARC:UPI0000173C42; UNIPARC:UPI0000173C43
J:Dubler, B.; van der Reer, M.
J. Biol. Chem. 262, 17724-17727, 1987
A:Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-
A:Reference number: S22254; MUID:88087065; PMID:3121603
A:Accession: S22254
A:Molecule type: protein
A:Residues: 2831-2832,'T',2834,'R',2836-2843,3002-3014 <DNB>
A:Cross-references: UNIPARC:UPI0000173C44; UNIPARC:UPI0000173C45
R:Tueb, J.; Tueb, B.
Biochim. Biophys. Acta 1171, 97-98, 1992
A:Title: The two splice variants of collagen XII share a common 5' end.
A:Reference number: S28811; MUID:93042014; PMID:1420368
A:Accession: S28811
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24,1189-1257,'S',1259-1263,'E',1265-1280 <TRU>
A:Cross-references: UNIPARC:UPI0000173C46; EMBL:X67327
C:Genetics:
A:Keywords: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
A:Cross-references: alternative splicing; cell binding; coiled coil; connective tissue; disulfide
C:Keywords: signal sequence #status predicted <SIG>
F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
F:24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted
F:24-105/Domains: fibronectin type III repeat predicted <IIIA>
F:24-301/Domains: von Willebrand factor type A repeat homology <FN3A>
F:333-425/Domains: IIB #status predicted <IIIB>
F:333-414/Domains: fibronectin type III repeat homology <FN3B>
F:437-601/Domains: von Willebrand factor type A repeat homology <VWA2>
F:623-1178/Domains: IIC #status predicted <IIIC>
F:630-711/Domains: fibronectin type III repeat homology <FN3C>
F:721-802/Domains: fibronectin type III repeat homology <FN3D>
F:812-895/Domains: fibronectin type III repeat homology <FN3E>
F:905-966/Domains: fibronectin type III repeat homology <FN3F>
F:995-1076/Domains: fibronectin type III repeat homology <FN3G>
F:1086-1169/Domains: fibronectin type III repeat homology <FN3H>
F:1137-1361/Domains: von Willebrand factor type A repeat homology <VWA3>
F:1384-2295/Domains: IIID #status predicted <IIID>
F:1384-1465/Domains: fibronectin type III repeat homology <FN3I>
F:1474-1557/Domains: fibronectin type III repeat homology <FN3J>
F:1566-1647/Domains: fibronectin type III repeat homology <FN3K>
F:1565-1738/Domains: fibronectin type III repeat homology <FN3L>
F:1766-1838/Domains: fibronectin type III repeat homology <FN3M>
F:1847-1928/Domains: fibronectin type III repeat homology <FN3N>
F:1937-2019/Domains: fibronectin type III repeat homology <FN3O>
F:2028-2110/Domains: fibronectin type III repeat homology <FN3P>
F:2119-2199/Domains: fibronectin type III repeat homology <FN3Q>
F:2207-2294/Domains: fibronectin type III repeat homology <FN3R>
F:2335-2490/Domains: von Willebrand factor type A repeat homology <VWA4>
F:2438-2440/Domains: cell adhesion #status predicted
F:2509-2750/Domains: IIV, homologous to NC4 domain of type IX collagen #status predicted
F:2751-2902/Domains: collagenous COL2 #status predicted <COL2>
F:2899-2901/Region: cell attachment (R-G-D) motif
F:2903-2945/Domains: non-collagenous NC2 #status predicted <NC2>
F:2946-3048/Domains: collagenous COL1 #status predicted <COL1>
F:3049-3124/Domains: non-collagenous NC1 #status predicted <NC1>
F:3124-2089,2093,1044,1512,1767,2210,2273,2532,2683/Binding site: hydroxyproline (Pro) #state
F:7280,2788,2836,2842,2860,2866,2869,3004,3007/Modified site: hydroxyproline (Pro) #state

Alignment Scores:	
Pred. No.:	2.97e-18
Score:	444.00
Percent Similarity:	47.2%
Best Local Similarity:	31.1%
Query Match:	18.7%
Length:	3124
Matches:	123
Conservative:	64
Mismatches:	169
Indels:	40

Db 763 GIUARGARGlnValThrValSerAlaAsncluarGserThrThrLeuArgAsnLeulle 782
QY 1099 CCGGACCGCCCTTACCTGTCGTACCGTGACCGCGCCGCTTCCGTCGCGCCGACGAGCGG 1158
Db 783 ProAspThrArgTyrGlnValSerValIleIleaglTufYrGlnSerGlyProGlyAsnAla 802
QY 1159 CTGTCCGCCAAGCGCTGCACAGCCCGACGCGCGCGCGCGCGCCCGCCACGC 1206
Db 803 LeuAsnGlyTyrAlaIuYrThr--AspGluValAlaGlyAsnProArg 817
RESULT 2
A:54849
collagen alpha 1 (VII) chain precursor - human
N:Alternate names: procollagen alpha 1 (VII) chain
C:Species: Homo sapiens (man)
C:Date: 04-Nov-1994 #sequence, revision 04-Nov-1994 #text change 09-Jul-2004
C:Accession: A54849; PH0844; S16316; I56328; A30296; I84686
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Ultio, J.
J. Biol. Chem. 269, 20256-20262, 1994
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha (VII)
A:Reference number: A54849; MUID:94327588; PMID:8051117
A:Accession: A54849
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-294 <CHR>
A:Cross-References: UNIPROT:Q02388; UNIPARC:UPI000017A138; GB:L02870; NID:9897124; PIDN:
R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A:Title: Molecular cloning and characterization of type VII collagen cDNA.
A:Reference number: PH0844; MUID:92231902; PMID:1567409
A:Accession: PH0844
A:Molecule type: mRNA
A:Residues: 'EFR', 340-475, 'RALSTASHSTLCMRATRWHPNCRGSHWTACEPCNRPASHPRAARAG', 524-528, 'C',
A:Cross-References: UNIPARC:UPI000017A139; DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:
A:Experimental source: keratinocyte
A:Note: the authors translated the codon ACC for residues 394 and 397 as Tyr
R:Parente, M.G.; Chung, L.C.; Rymaenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A:Reference number: S16316; MUID:91334380; PMID:1871109
A:Accession: S16316
A:Molecule type: mRNA
A:Residues: 815-892, 'E', 894-1439 <PAR>
A:Cross-References: UNIPARC:UPI000016A722; GB:M65188; GB:S49017; NID:8180914; PIDN:AAA98
A:Experimental source: keratinocyte
R:Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prietayanh, P.S.; Cook, M.E.; Wright, J.
J. Invest. Dermatol. 99, 631-636, 1992
A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot
A:Reference number: I56328; MUID:93107742; PMID:1469284
A:Accession: I56328
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'EFR', 372-517, 'DV', 520-540, 'W', 542-1255 <RBS>
A:Cross-References: UNIPARC:UPI000016B3AC; GB:S51236; NID:9262308; PIDN:AAB24637.1; PID:
R:Seitzler, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgess, R.E.
J. Biol. Chem. 264, 3822-3826, 1989
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase
A:Reference number: A30296; MUID:89139437; PMID:2537292
A:Accession: A30296
A:Molecule type: protein
A:Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041, '
A:Cross-References: UNIPARC:UPI000017A13A; UNIPARC:UPI000017A13B; UNIPARC:UPI000017A13C;
A:Note: two reported peptides cannot be reliably located
R:Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A:Reference number: I48103; MUID:93271985; PMID:8495916
A:Accession: I48103
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 2395-2871, 'S', 2873-2944 <RE2>
A:Cross-References: UNIPARC:UPI000016A724; GB:L06862; NID:9388713; PIDN:AAA89196.1; PID:

R.Christiano, A.M.; Rynaen, M.; Uitto, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A>Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser sub
A:Reference number: A55255; MUID:94224777; PMID:8170945
A:Contents: annotation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Genetic8:
A:Gene: GDB:COL7A1; EBR1; EBD1; EB
A:Cross-references: GDB:128750; OMIM:120120
A:Map position: 3p21.3-3p21.3
A:Map: defects in this gene can result in dominant and recessive dystrophic epidermoly
A>Note: there are 118 introns
C:Function:
A:Description: structural component of extracellular polymer associated with anchoring
C:Keywords: collod coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-2944/Product: collagen alpha 1(VI) chain #status predicted <MAT>
F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F:36-201/Domain: von Willebrand factor type A repeat homology <VWA1>
F:321-318/Domain: fibronectin type III repeat homology <FN1>
F:327-413/Domain: fibronectin type III repeat homology <FN2>
F:414-502/Domain: fibronectin type III repeat homology <FN3>
F:508-593/Domain: fibronectin type III repeat homology <FN4>
F:598-683/Domain: fibronectin type III repeat homology <FN5>
F:686-771/Domain: fibronectin type III repeat homology <FN6>
F:776-862/Domain: fibronectin type III repeat homology <FN7>
F:864-952/Domain: fibronectin type III repeat homology <FN8>
F:954-1045/Domain: fibronectin type III repeat homology <FN9>
F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>
F:1170-1172/Region: cell attachment (R-G-D) motif
F:1189-1253/Region: cysteine/proline-rich
F:1254-2783/Region: interrupted helical
F:1334-1336/Region: cell attachment (R-G-D) motif
F:2008-2010/Region: cell attachment (R-G-D) motif
F:2553-2555/Region: cell attachment (R-G-D) motif
F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F:2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F:337/786,1109/Binding site: carboxylate (asn) (covalent) #status predicted
F:2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status expe
F:2653,2631/Modified site: 5-hydroxylysine (lys) #status experimental
F:2653,2631/Binding site: carboxylate (lys) (covalent) #status experimental
F:2653,2802,2804/Dissulfide bonds: interchain #status predicted

Alignment Scores:
Pred. No.: 5,44e-15 Length: 2944
Score: 387.00 Matches: 137
Percent Similarity: 41.9% Conservative: 47
Best Local Similarity: 31.2% Mismatches: 205
Query Match: 16.3% Indels: 50
DB: 2 Gaps: 12

US-10-699-035A-5 (1-1254) x A54849 (1-2944)

QY 31 CTGAGCTTGGCGGCTGGCGGAGCGGAGCGGAGCGGAGTTCACACGATCAGCC 90
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MethrileuAArgleuValAlaAlaLeuCysAla-----GlyIleLeuAlaGluAla 18
QY 91 CCCCCG-----GGGAC 102
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
19 ProArGValAArgAlaGlnHisArgGluAArgValAlhCysThrAArgLeuTyrAlaAlaAsp 38
QY 103 CTGATGTTCTCGTGGACAGCTGACGAGCGAGCTCTTCACATGAGATTCCTCCGGGTTCCG 162
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
39 IleValIleLeuLeuLeuAlaGlySerSerSerIleLeuIleArgSerIleAlaValArg 58
QY 163 GAGTTTGGGGGAGCTGGTGGCTTCACATGCGCCCTGGGAGCAGGGAGCC-----CTGCGT 216
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
59 SerPheLeuGluGlyLeuValIleuProPheSerGlyIleAlaSerAlaGlnGlyValArg 78
QY 217 GCACGTCTGGTGCACGTCGGGACGTCGGCATACCCGAGTTCCCTTCGGCAGACAGC 276
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```
D 79 PheAlaThrValGIInTySerAspAspProAlaGThrGluPheGlyLeuAspAlaLeuGly 98
Q 277 TCGGGGTGAGGCGTCCCGAGATGCGGTGCTGCTTCCGCCAGCGATGGGTGACACCCAC 336
D 99 SerGIyGLIAspValIleArgAlaIleArgGluLeuSerTyrLeuGlyLysAsnThrArg 118
Q 337 ACTGGCTGGCGCGTGTCTATGCCAAGACAGCTGTTCGTGAGACATCAGGTGCCCG 396
D 119 ThrGIyAlaAlaIleLeuHISValAlaAspHISValPheLeuProGluLeu---AlaArg 137
Q 397 CCAGGGGTGCCCAAGTGTGGTGGTGGAGATGGCGGCTCCAGCGACCTGGCGGAC 456
D 138 ProGIyValIleProIyValCysIleLeuIleThrAspGlyLysSerGlnAspLeuValAsp 157
Q 457 CCCCCATGACAGAGCTCAAGACCTGGGCGTCAACCGTTCATGTCAGACCCGCCGA 516
D 158 ThrAlaAlaGlnArgLeuLysGlyGlnGlyValLysLeuPheAlaValGlyIleLysAsn 177
Q 517 GGCMACTTCTGGAGCTGTTCAGCCGCTGACGCCCTCGAGAGAGACACTGCACCTT 576
D 178 AlaAspProGluGluLeuLysArgValAlaAspGlnProThrSerAspPhePhe 197
Q 577 GTGAGAGTGTGATGACCTGCACATCATT-----GNC 606
D 198 -----ValAsnAspPheSerIleLeuArgThrLeuLeuProLeuValSerArgVal 215
Q 607 CAAGAGCTGAGGGGCTCATCTCGGAGTCG-----CCG 642
D 216 CysThrThrAlaGlyValIleProValIleThrArgProProAspAspSerThrSerAlaPro 235
Q 643 CAGGAGCTCCATGACCGAGATCAGCTCCAGGCGTTCGCGCTGCGCCACCCCTG 702
D 236 ArgAspLeuValLeuSerGlnProSerSerGlnSerLeuArgValGlnThrAlaAla 255
Q 703 CTGACCGGAGCTGGGCTACTATG-----CTGAGCTGGTGGCCAGCGCCGCG 756
D 256 SerGIyProValIleThrIyTyrIyValGlnTyrThrProLeuThrGlyLeuGlnPro 275
Q 757 GGGGCTGCAAGACGCCAGAG-----CTGCCAGGGAGCGCCAGCTGATCTGGGCG 810
D 276 LeuProSerGlnArgGlnGlnValAsnValProAlaGlyLysThrSerValArgLeuArg 295
Q 811 GGCCTGACCGCGGACCGAGATCGAGTGGCGCGCTGAGCTCAGTCCAACTGCCCTC 870
D 296 GlyLeuArgProLeuThrGluTyrGlnValIleValIleAlaLeuTyrAlaAsnSerIle 315
Q 871 CTGAGGCGCCAGATCCTGGGCGGAGCGGAGCGCGCCAGAGGCGCGGAGCGGCATC 930
D 316 GlyIyAlaValIleSerIyThrAlaArgThrAlaLeuGln---GlyProGlu---Leu 333
Q 931 GTCATCTCCAGCGCCGCGCGGAGCGCTCCGCGTGTGGAGTGGCCCGCGCTGGAGTCA 990
D 334 ThrIleGlnAsnThrThrAlaHisSerLeuLeuAlaIleThrArgSerValProGly--- 352
Q 991 GCCCGCGCGCTGGCTACCACTGACGTTGGGCGCGCTGGGCGGGGAGGCGGACGG 1050
D 353 -----AlaThrGlyTyrArgValIleThrTyrArgValLeuSerGlyGlyProThrGlnGln 370
Q 1051 GTGAGGTCGCCGCGCGGAGCACTGCACACAGCTGACGAGGCGCGCGCGGAGCGCG 1110
D 371 GlnGluLeuGlyIleProGlyGlnGlySerValLeuLeuArgAspLeuGluProGlyThrAsp 390
Q 1111 TACCTGTGACCGTGAACGCGCGCTTCCGCTCGGGCGCGGAGCGCGCTGTCGCGCAAG 1170
D 391 TyrGluValIleThrValSerThrLeuPheGlyArgSerValGlyProAlaThrSerLeuMet 410
Q 1171 GCGTGAACGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAC 1227
D 411 AlaArgThrAspAlaSerValGluGlnThrLeuArgProValIleLeuGlyProThr 429
```

RESULT 3
S31212
collagen alpha 1(XIV) chain precursor, short form - chicken

```
C/Species: Gallus gallus (chicken)
C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 15-Sep-2003
C/Accession: S31212
R/Waechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A/Title: Complete primary structure of chicken collagen XIV.
A/Reference number: S31211; MUID:93185668; PMID:8444186
A/Accession: S31212
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: mRNA
A/Residues: 1-1857 <MAE>
A/Cross-references: UNIPARC:UPI000017123C; EMBL:X70792; NID:9288874; PIDN:CA50063.1; PI
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C/Genetics:
A/Genes: Col14A1
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F.1-28/Domain: signal sequence #status predicted <SIG>
F.29-110/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>
F.29-110/Domain: fibronectin type III repeat homology <FN3A>
F.156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F.352-433/Domain: fibronectin type III repeat homology <FN3B>
F.442-525/Domain: fibronectin type III repeat homology <FN3C>
F.534-614/Domain: fibronectin type III repeat homology <FN3D>
F.623-707/Domain: fibronectin type III repeat homology <FN3E>
F.741-823/Domain: fibronectin type III repeat homology <FN3F>
F.832-914/Domain: fibronectin type III repeat homology <FN3G>
F.922-1009/Domain: fibronectin type III repeat homology <FN3H>
F.1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Alignment Scores:
Pred. No.: 1,1e-14 Length: 1857
Score: 382.00 Matches: 118
Percent Similarity: 43.8% Conservative: 61
Best Local Similarity: 28.9% Mismatches: 158
Query Match: 16.1% Indels: 72
DB: 2 Gaps: 13

US-10-699-035A-5 (1-1254) x S31212 (1-1857)
Q 79 CCAGATGACCGCCCGGAGGAGCTGATGTTCTGTGACAGCTCAGCGCTCTCT 138
D 154 ProAlaIleAla-----AspIleValIleLeuValAspGlySerThrSerIleGly 170
Q 139 CACTACAGAGTTCTCCGGGTTGCGGAGTTTGTGGGCGAGCTGGTGCCTCCATGCCCTG 198
D 171 ArgPheAsnPheArgLeuValArgLeuPheLeuGlnAsnLeuValSerAlaPheAsnVal 190
Q 199 GGCAGCGGGGCGCGTCCGCTGCGACGTCGTGGCGGCGGCGGCGCATACCGAGTTTC 258
D 191 GlySerGluTyrThrArgValGlyLeuAlaGlnTyrSerGlyAspProArgIleGluTyr 210
Q 259 CCTTCGCGCAGACAGCGCGGCTGGAGCTGCCAGATGCGGTCGCTTCTCCCGAG 318
D 211 HisLeuAsnAlaTyrGlyThrIyAspAlaValIleLeuAspAlaValArgAsnLeuProTyr 230
Q 319 CGCATGGGTGACACCCACACTGCGCTGGCGCTGTATGCCAAGAACAGCTGTTGCT 378
D 223 IySGlyGlyAsnThrLeuThrGlyLeuAlaLeuThrTyrIleLeuGlnAsnSerPheLys 250
Q 379 GAAAGATCAGGTGCGCGCGGAGGAGTGGCGGCGCAAGTCGTGTGGGTGACAGATGGCGCG 438
D 251 ProGluAlaGlyAlaArgProGlyValSerIyIleGlyIleLeuIleThrAspGlyLys 270
Q 439 TCCAGAGACCTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 498
D 271 SerGlnAspAspValIleProProAlaIyAsnLeuAlaGAspAlaGlyIleIleLeuPhe 290
Q 499 ATTGTCAAGACCGCGGAGGCACTTCTGAGCTGTCAAGCGGCTGACCGCTGCGC 558
D 291 AlaIleGlyValIyAsnAlaAspIleAsnGlyLeuLysGluIleAlaSerGluProAsp 310
Q 559 GAAAGCACCTGCACCTTGTG---GACGTGATGACCTGACATCATTTGCTCAA----- 609
```

Db 311 SerThrHisValIyrAsnValAlaAspPheAsnPhemeCAsnSerIleValGluGlyLeu 330
 QY 610 -----GAGCTGAGGGGCTCCATTCTC 630
 Db 331 ThrArgThrValCysSerArGValGluGluGlnLulysIleuIleValGlyThrIleAla 350
 QY 631 GCGATG-----CGCCGCGAGCAGCTTCATGCCAGAGATCAAGCTACGCTCCAGCGCTTCGCC 664
 Db 351 AlaSerIleuGlyAlaProThrAspLeuValThrSerAspIleThrAlaArgGlyPheArg 370
 QY 685 CTGGCGCTGGCCACCCCTGCTGAGCCAGAGCTCGGGC----- 720
 Db 371 ValSerTrp-----ThrHisSerProGlyLysValGluLysIyrArGValVal 386
 QY 721 TACTATGTGTGAGCTGTGTGCTCCAGCGCCAGCGGCTGCAAGACCCAGCAGCTG 780
 Db 387 TyrTyr-----ProThrArgGlyGlyGlnProGluGluValValVal 400
 QY 781 CCAGGAGACCCAGCAGCTGATCTGGCGCGCTCGACCCGAGACAGCAGTACAGCTG 840
 Db 401 AspGlySerSerSerThrAlaValLeuLysAsnLeuMetSerLeuThrGlyIle 420
 QY 841 GCGCTAGTG-----CCTGAGTCCAC 861
 Db 421 AlaValPheAlaIleTyrSerAsnAlaAlaSerGluGlyLeuArgGlyThrGluThrThr 440
 QY 862 GTGCGCTCTCTGAGCCCGCAGATCTGCGGGTGCAGCAGCGCAGAGAGCGCGGACCA 921
 Db 441 LeuAlaLeu-----PrometAlaSerAspLeu 449
 QY 922 GAGCGCATCTCATCTCCACAGCCCGCGCGCAGCCTCGCTGAGTTGGCCCGACGC 981
 Db 450 LysLeuTyrAspValSerHisSer-----SerMetArgAlaLysTrpAsn----- 464
 QY 982 CTGGCGTACGCGCGCGCTCGGCTCCACAGCTGACGTGCGCGCTCGCGCGCGG--- 1038
 Db 465 ---GlyValAlaIleAlaThrGlyIleTyrMetIleLeuIyrAlaProLeuThrGluGlyLeu 483
 QY 1039 GAGGCCAGCGGCTGAGGTGAGTGCAGCGCGCGCGCAGCTGACAGCAGCTGACG---GCG 1092
 Db 484 AlaIleAspGluLysGluIleIleGlyGluAlaSerThrGluLeuGluLeuAspGly 503
 QY 1093 CTGGCGCGCGGACCGCTACCTGCTGAGCAGCTGACCGCGCTTCGCTCGCGCGCGAG 1152
 Db 504 LeuLeuProAsnThrGlyIyrThrValThrValIyrAlaMetPheGlyGluGluAlaSer 523
 QY 1153 AGCGGCTGTCCCGCAAGCGCTGACG 1179
 Db 524 AspProLeuThrGlyGlnGluThrThr 532
 Db
 RESULT 4
 S78476
 collagen alpha 1(XIV) chain precursor, long form - chicken
 C/Species: Gallus gallus (chicken)
 C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 09-Jul-2004
 C/Accession: S78476, S31211
 R./Trueb, B.
 submitted to the EMBL Data Library, January 1993
 A./Reference number: S78476
 A./Accession: S78476
 A./Molecule type: mRNA
 A./Residues: 1-1888 <TRU>
 A./Cross-references: UNIPROT:P32018; UNIPARC:UPI0000126D31; EMBL:X70793; NID:G288872; PIR
 R./Maechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
 Eur. J. Biochem. 212, 483-490, 1993
 A./Title: Complete primary structure of chicken collagen XIV.
 A./Reference number: S31211, W01D:93185668; PMID:8444186
 A./Accession: S31211
 A./Status: Preliminary
 A./Molecule type: mRNA
 A./Residues: 1-416-1460-1811, 1843-1888 <WAB>
 A./Cross-references: UNIPARC:UPI00001773F4; UNIPARC:UPI00001773F5; EMBL:X70793
 C./Genetics:

A./Gene: COL1A1
 C./Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim
 F./1-28/Domain: signal sequence #status predicted <SIS>
 F./1-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <MAT>
 F./29-110/Domain: fibronectin type III repeat homology <FN3>
 F./156-320/Domain: von Willebrand factor type A repeat homology <WMA1>
 F./352-433/Domain: fibronectin type III repeat homology <FN3B>
 F./442-525/Domain: fibronectin type III repeat homology <FN3C>
 F./534-614/Domain: fibronectin type III repeat homology <FN3D>
 F./623-707/Domain: fibronectin type III repeat homology <FN3E>
 F./741-823/Domain: fibronectin type III repeat homology <FN3F>
 F./832-914/Domain: fibronectin type III repeat homology <FN3G>
 F./922-1009/Domain: fibronectin type III repeat homology <FN3H>
 F./1040-1205/Domain: von Willebrand factor type A repeat homology <WMA2>
 Alignment Scores:
 Pred. No.: 1,1e-14 Length: 1888
 Score: 382.00 Matches: 118
 Percent Similarity: 43.8% Conservative: 61
 Best Local Similarity: 28.9% Mismatches: 158
 Query Match: 16.1% Indels: 72
 DB: 2 Gaps: 13
 US-10-699-035A-5 (1-1254) x S78476 (1-1888)
 QY 79 CCAGATCAGCCCCCGAGGGACCTGATGTTCTCTGACAGCTCAGCAGCTCTCT 138
 Db 154 ProAlaIleAla-----AspIleValIleLeuValAspGlySerIyrSerIleGly 170
 QY 139 CACTACAGATTCTCCCGGTTGGAGATTGTGGGGCAGCTGCTGCTCCTCAGCCCTG 198
 Db 171 ArgPheAsnPheArgLeuValArgLeuPheLeuGluAsnLeuValSerIlePheAsnVal 190
 QY 199 GGCACGGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 258
 Db 191 GlySerGluLysThrArgValGlyLeuAlaGlnTyrSerGlyAspProArgIleGluTrp 210
 QY 259 CCTTCGCGCAGCAGCAGCTCGGCTGAGGCTGCCAGATGCGGCTGCTCTGCGCCAG 318
 Db 211 HisLeuAsnAlaTyrGlyThrLysAspAlaValLeuAspAlaValArgAsnLeuProTyr 230
 QY 319 CCGATGGTGACACCCACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 378
 Db 231 LysGlyGlyAsnThrLeuThrIleThrIleLeuAlaLeuThrIleLeuGluAsnSerPheLys 250
 QY 379 GAAGATCAGGTCGCGCGCAGCGGCGCCAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 438
 Db 251 ProGluAlaGlyAlaArgProGlyValSerLysIleGlyIleLeuIleThrAspGlyLys 270
 QY 439 TCCAGCGACCTCTGTGGGCCCCCCCATGCAAGAGCTCAAGACCTGGGCTCACCGTTC 498
 Db 271 SerGlnAspAspValIleProProAlaLysAsnLeuArgAspAlaGlyIleGluLeuPhe 290
 QY 499 ATTGTACAGACCGCGCGCAGCACTTCTCGAGCTGTGACCGCGCTGCTGCTGCTGCTG 558
 Db 291 AlaIleGlyValLysAsnAlaAspIleAsnGluLeuLysGluIleAlaSerIleuProAsp 310
 QY 559 GAGAGCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 609
 Db 311 SerThrHisValIyrAsnValAlaAspPheAsnPhemeIAsnSerIleValGluIleu 330
 QY 610 -----GAGCTGAGGGGCTCCATTCTC 630
 Db 331 ThrArgThrValCysSerArGValGluGluGlnLulysIleuIleValGlyThrIleAla 350
 QY 631 GCGATG-----CGCCGCGAGCAGCTTCATGCCAGAGATCAAGCTACGCTCCAGCGCTTCGCC 684
 Db 351 AlaSerIleuGlyAlaProThrAspLeuValThrSerAspIleThrAlaArgGlyPheArg 370
 QY 685 CTGGCGTGGCCACCCCTGCTGAGCCAGAGCTCGGGC----- 720
 Db 371 ValSerTrp-----ThrHisSerProGlyLysValGluLysIyrArGValVal 386

QY 721 TACTATGCTGAGAGTGGTCCAGCGCCAGCGCGGGGCTGCAAGACGACGCTG 780
Db |||||
387 TTTTyr-----ProthraArgylGlnProGluGluValValVal 400
QY 781 CCAGGAAACGACGAGCTGATCTGGCGCGCTGACCCGACGACGACTACGACGTG 840
Db |||||
401 AspGlySerSerSerThrAlaValLeuLysAsnLeuMetSerLeuThrGlnGlnIle 420
QY 841 GCGCTAGTGT-----CCTGAGTCCAAC 861
Db |||||
421 AlaValAlaPheAlaIleTyrSerAsnAlaIleSerGlnGluGlyLeuArgGlyThrGlnThr 440
QY 862 CTGCGCCCTCTTACGAGCCCAAGATCCGCGGGTCCGACCGCGCCAGAGAGCGCGGCA 921
Db |||||
441 LeuAlaLeu-----PrometAlaSerAspLeu 449
QY 922 GAGCGCATCGTATCTCCACGCGCGCGCGCGCGCGCTCGGCTGAGTGGGCGCCACAGG 981
Db |||||
450 LysLeuTyrAspValSerHisSer-----SerMetArgAlaLysTyrPasn----- 464
QY 982 CTGGGCTCACCGCGCGCGCTCGGCTACCACTGACAGTTCGCGCGCGCGCGG--- 1038
Db |||||
465 ---GlyValAlaGlyAlaThrGlyTyrMetIleLeuTyrAlaProLeuThrGlnGlyLeu 483
QY 1039 GAGCGCGACGCGGTGAGGTGCCCGCGCGCGCGCGCACTGACCACTGCTGACG-----GGC 1092
Db |||||
484 AlaAlaAspGluLysGlnIleLysIleGlyGluAlaSerThrGlnLeuGlnLeuAspGly 503
QY 1093 CTGGCGCGCGGACGCGCTACCTGAGTGGTACCGTACCGCGCGCTCGCGCGCGCGGAG 1152
Db |||||
504 LeuLeuProAsnThrGlnTyrThrValThrValTyrAlaMetPheGlyGluGluAlaSer 523
QY 1153 AGCGCGCTGTCCGCGCAAGGCTGACG 1179
Db |||||
524 AspProLeuThrGlyGlnGlnThrThr 532

RESULT 5
collagen alpha 1(XIV) chain precursor, short form 2 - chicken
N.Alternate names: undulin
C.Species: Gallus gallus (chicken)
C.Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C.Accession: A45974; S30085; S22916; S17035; S20833
R.Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gemmari, M.; Dublet, B.; Cancedda, R.; Lin
J. Biol. Chem. 268, 12177-12184, 1993
A>Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region
na.
A.Reference number: A45974; MUID:93280195; PMID:8505337
A.Accession: A45974
A.Status: preliminary
A.Molecule type: mRNA; protein
A.Residues: 1-1747 <GER>
A.Cross-references: UNIPROT:P32018; UNIPARC:UPI0000173C47
A.Experimental source: embryo skin
A>Note: sequence inconsistent with the nucleotide translation
R.Appe, S.S.
submitted to the EMBL Data Library, March 1992
A.Reference number: S30085
A.Accession: S30085
A.Molecule type: mRNA
A.Residues: 1472-1660 <APT>
A.Cross-references: UNIPARC:UPI00000006A2; EMBL:X65122; NID:g62871; PIDN:CAA46238.1; PID
R:Trueb, J.; Trueb, B.
Eur. J. Biochem. 207, 549-557, 1992
A>Title: Type XIV collagen is a variant of undulin.
A.Reference number: S22916; MUID:92339443; PMID:1339349
A.Accession: S22916
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 286-494, 'O', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>
A.Cross-references: UNIPARC:UPI0000173C48
R.Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsenmayer, T.F.; van der Rest, M.; Mayne,

Eur. J. Biochem. 201, 333-338, 1991
A>Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens
A.Reference number: S17035; MUID:92037585; PMID:1935930
A.Accession: S17035
A.Molecule type: mRNA
A.Residues: 1472-1659 <GOR1>
A.Cross-references: UNIPARC:UPI0000173C49
A.Accession: S20833
A.Molecule type: protein
A.Residues: 1551-1570; 1593-1599; 1639-1667 <GOR2>
A.Cross-references: UNIPARC:UPI0000173C49
A.Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F:40-204/Domain: von Willebrand factor type A repeat homology <VWA1>
F:236-317/Domain: fibronectin type III repeat homology <FN3A>
F:326-409/Domain: fibronectin type III repeat homology <FN3B>
F:418-498/Domain: fibronectin type III repeat homology <FN3C>
F:507-591/Domain: fibronectin type III repeat homology <FN3D>
F:625-707/Domain: fibronectin type III repeat homology <FN3E>
F:716-798/Domain: fibronectin type III repeat homology <FN3F>
F:806-893/Domain: fibronectin type III repeat homology <FN3G>
F:924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>
F:1111-1352/Domain: non-collagenous NC4 #status predicted <NC4>
F:1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>
F:1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Alignment Scores:
Pred. No.: 4,71e-14 Length: 1747
Score: 371.00 Matches: 116
Percent Similarity: 43.54 Conservative: 62
Best Local Similarity: 28.48 Mismatches: 159
Query Match: 15.64 Indels: 72
DB: 2 Gaps: 13

US-10-699-035A-5 (1-1254) x A45974 (1-1747)

QY 79 CGAGATGACGCCCCCGAGGGAGACCTGATGTTCTCTGTCGACAGCTACGACGCTCTCT 138
Db |||||
38 ProAlaIleAla-----AspIleValIleValAlaPheGlySerTyrSerIleGly 54
QY 139 CACTACGAGTTCTCCCGGTTTCGAGGATTTGGGGGACGCTGTCGCTCCACTGCGCCCTG 198
Db |||||
55 ArgPheAsnPheArgLeuValAlaArgLeuPheLeuGlnLeuValSerArgPheAsnVal 74
QY 199 GGCACCGGGGCCCTGTCGACAGTGTGTGACAGTGGCGACAGTGGCCATPACCGAGTTTC 258
Db |||||
75 GlySerGluLysThrArgValGlyLeuAlaGlnTyrSerGlyAspProArgIleGluThr 94
QY 259 CCCTTGGCGACGACAGCTCGGTCGAGGCTGCCAGATGGGTCGCTTCTTCCGACG 318
Db |||||
95 HistLeuAsnAlaTyrGlyThrLysAspAlaValLeuAspAlaAspArgAsnLeuProTyr 114
QY 319 CGCATGGGTGACACCACTGAGCTGCGCGCTGCTGCTGCAAGACAGCTGTTTCT 378
Db |||||
115 LysGlyGlyAsnThrLeuThrGlyLeuAlaLeuThrTyrIleLeuGlnAsnCysPheLys 134
QY 379 GAAGCATGAGTCCCGCGCAGGGGTGCCAAGTGTGTCGTGTGGTGACAGATGGCGGC 438
Db |||||
135 ProGluAlaGlyAlaArgProGlyValSerLysIleGlyIleLeuIleThrAspGlyLys 154
QY 439 TCCAGCGACCTGTGGGCCCCCATGCAAGCTCAAGACCTGGGCGCTCACCGCTTTC 498
Db |||||
155 SerGlnAspAspValIleProProAlaLysAsnLeuArgGluAlaGlyIleGluLeuPhe 174
QY 499 ATTGCACACCGCGCGGAGCACTTCTGAGCGTGTGACGCGCTCGCTCACCGCTCC 558
Db |||||
175 AlaIleGlyValLysAsnAlaAspIleAsnGlnLeuLysGluIleAlaSerGluProAsp 194
QY 559 GAGAAGCATGTCACCTTTGTG---GACGTGATGACCTGACATCATTTGTCCA----- 609
Db |||||
195 SerThrIleValTyrAsnValAlaAspPheAsnPheMetAsnSerIleValIleGluLeu 214
QY 610 -----GAGCTGAGGGGCTCCATTTCTC 630
Db |||||

Db 215 ThrArgThrValCysSerArgValGluGluGlnGluLysGluIleVal 234
 QY 631 GCGATG-----CGCGCGCAGCATCTCCATGCCAGGAGATCACTCCAGCGCTTCGCG 684
 Db 235 AlaSerIleuGluValProThrAspLeuValThrSerAspIleThrAlaArgGlyPheArg 254
 QY 685 CTGGCCTGGCCACCCCTGCTGACCGCAGACTCGGCG----- 720
 Db 255 ValSerTrp-----ThrHisSerProGlyLysValGluValGlyValVal 270
 QY 721 TACTATGTGTGAGCTGTGTGCGCAGCGCGCGCGGCTGCAAGACCGCAGCAGCTG 780
 Db 271 TyrTrp-----ProThrArgGlyGlyGlnProGluGluValVal 284
 QY 781 CCAGGAGACCGCAGCATGATCTGGCGCGCGCTGACCGCGCAGCATGACGATG 840
 Db 285 AspGlySerSerSerThrAlaValLeuLysAsnLeuMetSerLeuThrGluValGlnIle 304
 QY 841 GCGCTAGTG-----CTTAGTCCAC 861
 Db 305 AlaValPheAlaIleTyrSerAsnAlaAlaSerGluGlyLeuArgGlyThrGluThr 324
 QY 862 GTGCGCTCTCTGAGCGCCAGATCTGCGGTTGGCGCAGCGCGCAGAGCGCGGCGCA 921
 Db 325 LeuAlaLeu-----ProMetAlaSerAspLeu 333
 QY 922 GAGCGCATGCTATCTCCACGCGCGCGCGCGCAGCGCTCCGCTGAGTTGGCGCGCAG 981
 Db 334 LysLeuTyrAspAlaSerHisSer-----SerMetArgAlaLysTrpAsn----- 348
 QY 982 CTGGGCTCAGCGCGCGCGCTGCTACACGTCGACGTTGCGGCGCTGCGGCGCGG-- 1038
 Db 349 ---GlyValAlaGlyAlaThrGlyTyrMetIleLeuTyrAlaProLeuThrGluGlyLeu 367
 QY 1039 GAGCGCGCAGCGGCGTGAAGCGCGCGCGCGCGCAGCTGCACGCTGCAG-----GGC 1092
 Db 368 AlaAlaAspGluLysGluIleTyrIleGlyGluAlaSerThrGluLeuGluAspGly 387
 QY 1093 CTGGCGCGCGCGCAGCGCTGCTACGTCGACGCGCTCCGCTGCGCGCGCGCAG 1152
 Db 388 LeuLeuProAsnThrGluTyrThrValThrValTyrAlaMetPheGlyGluGluAlaSer 407
 QY 1153 AGCGCGCTGTCCGCGCAAGCGCTGCAGC 1179
 Db 408 AspProLeuThrGlyGlnGluThrThr 416
 RESULT 6
 151027
 Type XII collagen alpha-1 chain - eastern newt (fragment)
 C:Species: Notoptthalmus viridescens, Triturus viridescens (eastern newt)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
 C:Accession: 151027
 R:Wei, Y.; Yang, E.V.; Klatt, K.P.; Tassava, R.A.
 Dev. Biol. 168, 503-513, 1995
 A:Title: Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII collagen
 A:Reference number: 151027; PMID:95246925; PMID:7729585
 A:Accession: 151027
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-929 <MBI>
 A:Cross-references: UNIPROT:Q91145; UNIPARC:UPI0000126D2E; EMBL:U19494; NID:9632647; PII
 F:155-236/Domain: fibronectin type III repeat homology <3PR>
 F:631-795/Domain: von Willebrand factor type A repeat homology <VMA3>
 Alignment Scores:
 Pred. No.: 2.16e-08 Length: 929
 Score: 272.50 Matches: 89
 Percent Similarity: 38.2% Conservative: 49
 Best Local Similarity: 24.7% Mismatches: 130
 Query Match: 11.4% Indels: 93
 DB: 2 Gaps: 9
 US-10-699-035a-5 (1-1254) x 151027 (1-929)

QY 64 GCGAGCGCGGTCA---CCAGCATCAGCC-----CCCGAGGGGAC 102
 Db 614 AlaProArgSerProIleProSerSerGlyLeuAspCysThrThrLysAlaGlnAlaAsp 633
 QY 103 CTGATGTTCTGCTGAGCAGCTCAGCAGCGGTCTTCACATCAGATTTCTCCGGTTCCG 162
 Db 634 IleValLeuLeuValAspGlySerTrpSerIleGlyArgProAsnPheLysIleValArg 653
 QY 163 GAGTTGTGGGAGCGCTGAGCGCTCCAGCCCTCCGCGCGCGCGCGCGCGCGT 222
 Db 654 AsnPheIleSerArgValValGluValPheAspIleGlySerAspArgValGlnIleAla 673
 QY 223 CTGTCACGTTGGCGCAGTCGCGCATACACGAGTTCCCTTCGCGCAGCAGCATCGGTT 282
 Db 674 ValSerGlnTyrSerGlyAspProArgThrGluTrpGlnLeuAsnThrHisLysThrLys 693
 QY 283 GAGCGTCCCGCAGATGCGCGTGGCTTCTGCGCCACGCGATGCGTACACCCACATCGCG 342
 Db 694 LysSerLeuMetAspAlaValAlaAsnLeuProTyrLysGlyLysAsnThrAsnThrGly 713
 QY 343 CTGGCGCTGCTATATCCAAAGAAAGAGCTGTTGTAAGCATCAGGTGCGCGCGCG 402
 Db 714 SerAlaLeuLysPheIleLeuGluAsnAsnPheArgProGlyValGlyMetArgGluLys 733
 QY 403 GTGCGCAAGTCTGTGCTGGGTGACAGATGGCGGCTCCAGCGACCTGTGGCGCGCGCG 462
 Db 734 AlaArgLysIleAlaIleLeuLeuThrAspGlyLysSerGlnAspIleValAlaPro 753
 QY 463 ATGCAGAGACTCAAGACCTGGCGCTCACCGTCTTCAATGTCAGCACCGCGCGCAGAAC 522
 Db 754 SerLysArgTyrAlaAspGluGlyIleGluLeuTyrAlaValAlaGlyLysAsnAlaAsp 773
 QY 523 TTCCTGAGCTGTGACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
 Db 774 GluAsnGluLeuLysGluIleAlaSerAspProAspGlu-----LeuTyrMetTyrAsn 791
 QY 583 GTGATGATCGTCGACATCTTTCACAAAGCTGAGGGGCTTCATCTCGGATGCGCGCG 642
 Db 792 ValAlaAspPheSerLeuLeu----- 798
 QY 643 CAGCAGCTCCATGCCAGAGATCACGTCACGCGGCTTCGCGCTGCGCGCACCCCTG 702
 Db 799 -----ThrAsnIleValAsnAsp----- 804
 QY 703 CTGACCGCAGACTCGGGCTACTATGTCGTGAGCTGTGCGCGCGCGCGCGCGGCT 762
 Db 805 -----LeuThrGluAsnValCysAsnSerValLysGly--- 815
 QY 763 GCMAAGCGCAGCAGCTGCGCAGGAAAGCGCAGCAGCTGATCTGCGCGCGCTCGACCG 822
 Db 816 -----ProGly-----GlyLeuAsnPro 821
 QY 823 GACACGACTCAGCAGCTGGCGCTAGTCCAGTCCAGCTGCGCTTCGAGGCGCGCG 882
 Db 821 ----- 821
 QY 883 ATCTGCGGAGTGCGCAGCGCGCGCAGAGAGCGCGCGCAGCGCATTCATCTCCAC 942
 Db 822 -----ProSerAsnLeuValThrSerGlu 829
 QY 943 GCGCGCGCGCAGCTCCGCGTGAATGGGCGCGCGCGCGCTGAGCGCGCGCGCTC 1002
 Db 830 ProThrProArgSerPheArgValThrTrpValProPro-----SerGlnSerValGlu 847
 QY 1003 GCGTACCATGTCGATGTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
 Db 848 ArgPheLysValGluTyrTyrProValAlaGlyArgProGlnGluValTyrValArg 867
 QY 1063 GCGCGCGCGCAATCGCAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACCTGAGAC 1122
 Db 868 GlyThrGlnThrThrThrValLeuValGlyLeuLysProGluThrGluTyrTyrValAsn 887

Score: 248.50 Matches: 66
 Percent Similarity: 49.4% Conservative: 82
 Best Local Similarity: 37.1% Mismatches: 23
 Query Match: 10.4% Indels: 7
 DB: 2 Gaps: 2

US-10-699-035A-5 (1-1254) x A33809 (1-493)

```

QY      73  GGTCCACACGATACAGCCCCCGAGGG-----GACCTGATGTTCTCTG 114
       257  GilylsthCySerSerAlaCySerSerGlyGlySerGlySerAlaLeuAspLeuValPheLeu 276
QY      115  CTGAACAGCTCAGCAGCGCTCTCTCACTACAGATTCTCCCGGATTCCGAGATTGTGGG 174
       277  IleAspGlySerLySerSerValArgProGluAspPheGluLeuValLySerPheIleAsn 296
QY      175  CAGCGTGGCTCCACTGCCCTGGGACACGGGGCCCTGCGTCCAGTGTGTCACGTG 234
       297  GlnIleValGluSerLeuIleValSerGluLySGlnAlaGlnAlaGlyLeuValGlnTyr 316
QY      235  GGCAGTCGGCCATACACCGAGTTCCCTTCGGCCAGACAGCTCGGATGAGCTGCCAG 294
       317  SerSerSerValArgGlnIleuPheProLeuGlyGlnPheLyAsnLySerAspIleLyS 336
QY      295  GATCGCGTCCGCTTCTTCCCAAGCGCATGGGTGACACCCACACTGCGCTGGCGCTGTC 354
       337  AlaAlaValLySerMetAlaTyrMetGluLySGlyThrMetThrGlyGlnAlaLeuLyS 356
QY      355  TATGCCAAGGAAGAGTGTTCCTGCTGAAGATCAGGCGCCGACGGGGTGGCCAAAGTG 414
       357  TyrLeuValAspSerSerPheSerIleAlaAsnGlyAlaArgProGlyValProLyVal 376
QY      415  CTGTGGTGGGTGACAGATGCGGCTCCAGCGACCCCTGTGAGGCCCCCATGACGAGCTC 474
       377  GlyIleValPheThrAspGlyArgSerGlnAspTyrIleThrAspAlaIleLySAsla 396
QY      475  AAGGACCTGGGCGTCAACCGTGTTCATTGTTCAGACACCGCCGAGGCAACTTCTCGAGCTG 534
       397  LyAspLeuGlyPheArgMetPheAlaValGlyValGlyAsnAlaValGluAspGlyLeu 416
QY      535  TCAGCGCGTCCGCTCAGCGCTGCCGAGACCTTCGATCTTTGTGACGTGGAT 588
       417  ArgGluIleAlaSerGluProValAlaGluHis--TyrPheTyrThrAlaAsp 433
  
```

RESULT 10

J00405

hypothetical 119.5K protein (uvra region) - Micrococcus luteus

NrAlternate names: ORF 1 protein

C:Species: Micrococcus luteus; Micrococcus lysodeikticus

C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000

C:Accession: J00405

R:Shiota, S.; Nakayama, H.

Mol. Gen. Genet. 217, 332-340, 1989

A:Title: Micrococcus luteus homolog of the Escherichia coli uvra gene: identification of

A:Reference number: S04781; MUID:89364717; PMID:2549377

A:Accession: J00405

A:Molecule type: DNA

A:Residues: 1-1106 <SH1>

A:Cross-references: UNIPARC:UPI0000177394; EMBL:X15867

A:Note: this reading frame extends between two stop codons and does not begin with a sta

A:Note: the gene encoding this protein overlaps uvra gene

C:Superfamily: collagen alpha 1(1) chain; fibrillar collagen carboxyl-terminal homology;

Alignment Scores:

Pred. No.: 6.53e-07 Length: 1106
 Score: 246.50 Matches: 157
 Percent Similarity: 31.1% Conservative: 19
 Best Local Similarity: 27.7% Mismatches: 183
 Query Match: 10.4% Indels: 207
 DB: 2 Gaps: 29

US-10-699-035A-5 (1-1254) x J00405 (1-1106)

```

QY      24  CCTGGCCTTGAGCTTGGCGGCTGGCGCTGGCCGAGACGGCGCC-----GGAGCG 71
       569  ProGlyPro-----AlaGlyValProAlaGlyArgProArgValProGlnPro 585
QY      72  CCGTCCACAGATCAGCGCC-----CGAGGGGACCTGATGTTCTGTGTGACAGCTC 125
       566  ArgAlaAlaGlyArgHisProLeuArgArgArgGlyProAlaHisArgProGlyHisThr 605
QY      126  AGCCAG---CGTCTCTCACTACGAGTTCTC---CGGGTTCCGGAGTTTGTGGGCACT 179
       606  AspArgLeuArgAlaGlyArgArgProLeuArgArgProArgArgAlaValHisArgProAla 625
QY      180  GGTGCTCACTACCTCCCTGGGACACGGGGCCCTGCGTCCAGTCTGTGTGACAGT----- 233
       626  ProAlaGlyGlnProProHisArgArgProProAlaProAlaGlyProArgGlnHis 645
QY      233  ----- 233
       646  ProHisArgArgArgAlaArgArgGlyHisAspArgArgGlyGlyLeuAspArgGlyHis 665
QY      233  ----- 233
       666  ArgProSerArgGlyArgValArgArgArgGlyArgAlaLeuGlyLeuProGlyLySer 685
QY      234  ---GGGAGTGGCCATA-----CAGGAGTTCCCTTCGGCCAGCAGCTCGGATGA 284
       686  GlnGlyGlnHisAlaValArgHisArgArgArgProLeuArgPro----- 700
QY      285  GCGTCCGCA-----GAGTGGGTGCGTCTTCTGCTCCAGCGCATGGGTGACAC--- 332
       701  -----ProLeuHisArgGlyAlaGlyAlaAlaSerArgProGlyGlyGlyAlaArgAla 718
QY      333  -----CSACATGCGCTGGC-----GCTGCTTATGCGCAAGAACAGCTGT 374
       719  AspGlyProArgArgProGlyGlyGlnProGlyGlyArgGlyArgGlyProGly----- 735
QY      375  TGCTAAGCATAGGTGCGCGGCGTCCGAGGGGTGCGCCAAAGTCTGTGTGGGTGACAGATGG 434
       736  -----ProAlaArgGlyProHis-----GlyArgAspGlyArgVal 747
QY      435  CGGCTCCAG-----CGACCTGT-----GGGCCCCC 461
       748  ArgLeuArgGlnValHisAlaSerGlnArgSerProLeuGlnGlyProGlyGlnProAla 767
QY      444  -----CGACCTGT-----GGGCCCCC 461
       768  GlnArgArgGlnAlaArgAlaArgProAlaProAlaArgGlyGlyThrArgAlaProGly 787
QY      462  CATGCAGAGACTCAGAGACTGGGCGTACCGCTGTTCATTGTTCAGACACCGCGGAGGCA 521
       788  GlnGlyGlyProArgGlyProGluProHisArgAlaHisAlaThrLeuGlnProArgHis 807
QY      522  CTTCCTGAGAGCTGTCAGCGCGCTGCCCTCAGCGCCCTCCAGAAAGCACTGCACTTTGTGA 581
       808  LeuHisGlyArgValArgArgArgProGlnAlaLeuArgGlyAspArgProArgGlyGlyGly 827
QY      582  CGTGATGACCTGCA-----CATCATTTGTCAGAGCT---GAGGGGCTC 623
       828  ProGlyLeuProAlaGlyProValLeuLeuGlnHisGlnGlyArgAlaLeuArgGlyVal 847
QY      624  CATTTCGGGATCGGCGC----- 641
       848  ArgGlyArgArgHisAlaGluAspArgArgGlyLeuProAlaGlyArgLeuArgAlaVal 867
QY      641  ----- 641
       868  ArgGlyValProArgGlyProValGlnProGlyAspAlaArgGlyHisLeuGlnGlyGln 887
QY      642  -----GCAGCAGCTCCATGCGCACGAGATCAC-----GTCCAGCGGCTT 680
       888  GlnHisArgArgGlyProArgHisAlaLeuArgArgGlyGlyArgGlyLeuLeuGlnArgVal 907
QY      681  CCGCTCGGCTGGCCACCCCTGCTGAC-----CGCAGACTGGGCTACTATGTGCT 731
  
```

```

Db      908 HISPProHISLeuAlaValProGIYHISAlaArgArgArgSerGlyLeuArgProSer 927
      732 GGA-----GCTGATGCCAGCGCCAGCGGGGGCTGC 764
      928 GILProAlaArgHISHisAlaLeuGILArgArgGILProAlaArgGILAlaGILGly--- 946
QY      765 AAGACCGCAGACCTCCGACGAGAAAGCCACGACTGAGTGGCGCGCTCGACCCGGA 824
      947 ArgAlaProGILLeuGILArgProHISHisLeuArgAlaGILAlaAspHISGIL 966
QY      825 CACGAGCTAGAGTGGCGCTAGTGGCTGAGTCCACAGTGGCTCTTGAGGCCCCAGAT 884
      967 ValAlaLeuArgArg-----HISProGILAlaProAla 977
QY      885 CTGCGGAGTCCGACGCGGACGAGAGCGCGGCGCCAGACGAGCTCATCTCCACGC 944
      978 ArgAlaSerValProArgGILGILGILGILHISGILAlaHISHisArg----- 993
QY      945 CCGGCGCGGACGCTCCGCGT-----GAGTTGGGCGCCAGCGCTGGGCTCAGCCGCGC 998
      994 -----AlaGILProArgArgAspGILArgGILProArgAspArgProArgProGIL 1011
QY      999 GCTCGGCTACCACTGACGATTCGGCGCGCTGCGGCGGCGGAGCGCAGCGGCTGAGGT 1058
      1012 GILArgLeuArgArg-----ArgHISAspArgGILHISGILHISAlaGILGILGILArg 1029
QY      1059 GCCCGCGGCGCGGACACTGACACGACGCTGCGGCGCGGCGGACCGCTACCTGCT 1118
      1030 AlaArgArgArgGILProHIS-----GILPro-ValProArgGILAlaAlaArgVal 1046
QY      1119 GACCGTAGCCGCGCGCTTCGCTCGGCGCGGCGCGGACGCGCTGCTCCGACGAGCCCTGAC 1178
      1046 IAlaTrpThr-ProProThrGILAspGILAspGILArgGILGILGILGILGILGILGIL 1065
QY      1179 GCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1238
      1065 rgrProCysAlaGILAla-----ProProArgCysCysGILSerp 1078
QY      1239 CAGCCGTGAGCCGT 1252
      1078 rAlaValProArg 1082

RESULT 11
cartilage matrix protein precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C/Accession: S66522
R/Ascoli, A.; Hauser, N.; Studer, D.; Paulsson, M.; Hirtl, L.; Bosze, Z.
Eur. J. Biochem. 236, 970-977, 1996
A/Title: Cloning, sequencing and expression analysis of mouse cartilage matrix protein C
A/Reference number: S66522; MUID:96270751; PMID:8665920
A/Accession: S66522
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-500 <ASZ>
A/Cross-references: UNIPROT:P51942; UNIPARC:UPI000002480; EMBL:U35035; NID:G1163178; PI
C/Genetics:
A/Genes: CNP
C/Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe
F/1-29/Domain: signal sequence #status predicted <SIG>
F/30-500/Product: cartilage matrix protein #status predicted <MAT>
F/43-210/Domain: von Willebrand factor type A repeat homology <WMA1>
F/231-266/Domain: EGF homology <EGF>
F/277-441/Domain: von Willebrand factor type A repeat homology <WMA2>

Alignment Scores:
Pred. No.: 1.05e-06 Length: 500
Score: 243.50 Matches: 60
Percent Similarity: 50.9% Conservative: 29
Best Local Similarity: 34.3% Mismatches: 75
Query Match: 10.2% Indels: 11

```

```

DB:      2 Gaps:      2
US-10-699-035A-5 (1-1254) x S66522 (1-500)
QY      91 CCGCGAGG-----GACGTGATGTTCTCGTGCAGACTCA 126
      34 ProArgGILHISLeuCysArgThrArgProThrAspLeuValPheValAlaPheSer 53
QY      127 GCCAGCGTCTCATACGATGTTCTCCGAGTTCCGAGTTTGTGGAGGACGTTGAGCT 186
      54 ArgSerValArgProValGILPheGILuValValValPheLeuSerGILValIleGIL 73
QY      187 CCACTGCCCTTGCGGACCGCGGCGCTGCTGCCAGTGTGTGACGCTGGGCACTCGGCA 246
      74 SerLeuAspValGILProAspAlaThrArgValGILLeuValAsnTyraIaSerThrVal 93
QY      247 TACACGAGTTCCTTCGCGGACGACACTGCGGTTGAGGCTGCCAGATGCGGCGCT 306
      94 LysProGILPheProLeuArgAlaHisGILSerLysAlaSerLeuLeuGILAlaValArg 113
QY      307 GCTTTCGCCAGCGCATGGGTGACACCCACACTGAGCTGCGCTGCTATGCCAAGAA 366
      114 ArgIleGILProLeuSerThrGILyThrMetThrGILLeuAlaLeuGILHISAlaIleThr 133
QY      367 CAGCTGTTTCTGACATCAGTCCGCG-----CGAGGGTCCCAAGTCTG 417
      134 LysAlaLeuSerAspAlaGILGILGILGILArgAlaArgSerProAspIleSerLysValVal 153
QY      418 GTGTGGGTACAGATGGCGGCTCCAGGACCTGTGTGGCGCCCGCCCGCTGACAGAGCTCAAG 477
      154 IleValIleThrAspIleArgProGILAspSerValArgAspValSerGILuArgAlaArg 173
QY      478 GACCTGCGCTCACCGTGTTCATTTGACACCGCGCGGAGGACACTTCTGAGCTGTCA 537
      174 AlaSerGILyIleGILLeuPheAlaIleGILLeuGILArgValAspLysAlaThrLeuArg 193
QY      538 GCCGCTGCTCAGCCCTCGCGGAGAACACTGCACTTGTGTGAC 582
      194 GILHISAlaSerGILProGILAspGILHISValAspTyraValGILu 208

RESULT 12
QOBB3
BHLFI protein - human herpesvirus 4 (strain B95-8)
C/Species: human herpesvirus 4, Epstein-Barr virus
C/Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 09-Jul-2004
C/Accession: A03742
R/Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr viru
A/Reference number: A93065; MUID:85035713; PMID:6092825
A/Accession: A03742
A/Molecule type: DNA
A/Residues: 1-660 <BAN>
A/Cross-references: UNIPROT:P03181; UNIPARC:UPI000000CDDC
R/Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.;
Nature 310, 207-211, 1984
A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A/Reference number: A03794; MUID:84270667; PMID:6087149
A/Contents: annotation; protein coding region
C/Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-5
C/Superfamily: human herpesvirus 4 BHLFI protein

Alignment Scores:
Pred. No.: 1.02e-06 Length: 660
Score: 243.50 Matches: 150
Percent Similarity: 32.8% Conservative: 25
Best Local Similarity: 28.1% Mismatches: 183
Query Match: 10.2% Indels: 175
DB:      1 Gaps:      34
US-10-699-035A-5 (1-1254) x QOBB3 (1-660)
QY      1237 CGGTCCCGGCGTGGGCGCGCGGCGCGGCGCGGCGCTGGCG-----GCGGCGCGGCGC 1187

```



```

Db      890 AlaProAspThrArgProAlaProGlySerThr---AlaProProAlaHisGlyValThr 908
QY      307 CACGACCCGCAT-----CCTGGCAGACCTCACCCGAGCGTGTCTGGCCGAAG 260
Db      909 SerAlaProAspThrArgProAlaProGlySerThrAlaProProAla-----His 925
QY      259 GGAACCTCGGTATGATGCGCAGCTGCCACGACTGCAGCGACGACGCGCCGG--- 203
Db      926 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 945
QY      202 -----TGCCAGGGGCGAGTGAAGCCACACGCT--- 176
Db      946 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 965
QY      175 -----GCCCCCAAACT---CCGAAACCCGGGAGAACTGCTGTAGAGACGC 131
Db      966 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 985
QY      130 TGGCTGAGC-----TGTCCAGCAGAGAAATCATCAGTCCCTCGGG 92
Db      986 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 1005
QY      91 GGGCTGATCTGTGTC---GACCGCGCTCCGCGCGCTCCGCGCCAGCGCCAGCCGA--- 38
Db      1006 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 1025
QY      37 -----AGCTCAGGCGCCAGCGCGCTCCAGCGAGACA 2
Db      1026 GlyValThrSerAlaProAspThrArgProAlaProGlySer 1039

```

RESULT 14
A37979
cartilage matrix protein precursor - human

C:Species: Homo sapiens (man)
C>Date: 12-Jul-1991 #sequence_revision 12-Jul-1991 #text_change 09-Jul-2004
C:Accession: A37979; B37979
R:Jenkins, R.N.; Osborne-Lawrence, S.L.; Sinclair, A.K.; Eddy Jr., R.L.; Byers, M.G.; SH J. Biol. Chem. 265, 19624-19631, 1990
A>Title: Structure and chromosomal location of the human gene encoding cartilage matrix A:Reference number: A37979; MUID:91060568; PMID:2246248
A:Accession: A37979
A:Molecule type: DNA
A:Residues: 1-496 <JEN>
A:Cross-references: UNIPROT:P21941; UNIPARC:UPI000004F1ED; GB:J05667
A:Accession: B37979
A:Molecule type: mRNA
A:Residues: 157-290 'L', 292-496 <JEN>
A:Cross-references: UNIPARC:UPI000016A6E8; GB:M55683; GB:J05666; GB:J05667; NID:g180651;
C:Genetics:
A:Gene: GDB:CRIM
A:Cross-references: GDB:127280; OMIM:115437
A:Map position: 1p35-1p35
A:Introns: 32/1, 147/3; 222/1; 264/1; 403/1; 454/1; 481/1
C:Complex: homotrimer
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repeat
C:Keywords: glycoprotein; homotrimer
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-496/Product: cartilage matrix protein #status predicted <MAT>
F:39-206/Domain: von Willebrand factor type A repeat homology <WMA1>
F:227-262/Domain: EGF homology <EGF>
F:273-437/Domain: von Willebrand factor type A repeat homology <WMA2>
F:76,344/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:221-238,234-247,249-262/Disulfide bonds: #status predicted

Alignment Scores:

Pred. No.: 3,91e-06 Length: 496
Score: 233.50 Matches: 62
Percent Similarity: 50.3% Conservative: 27
Best Local Similarity: 35.0% Mismatches: 87
Query Match: 9.8% Indels: 1
DB: 2 Gaps: 1

US-10-699-035a-5 (1-1254) x A37979 (1-496)

```

QY      100 GACCTGATGTTCTCTGCTGAGACAGCTACAGCGCTCTCTCACTACGACTTCTCCGGGTT 159
Db      275 AspLeuValPheLeuIleAspGlySerIleSerValAArgProGluAsnPhelGluLeuVal 294
QY      160 CGGAGTTTGTGGGACAGCTGTGCTCCACTGCCCCCTGGGACCGGGGCTGTGCTGCC 219
Db      295 LysLysPheIleSerGlnIleValAspThrLeuAspValIleAspLysLeuAlaGlnVal 314
QY      220 AGCTGTGTCACGTGGGACGTGGCCATVACCCGAGTCCCTTCCCTGGCAGCAGACCTCG 279
Db      315 GlyLeuValGlnTyrSerSerSerValArgGlnGluPheProLeuGlyArgPheIleThr 334
QY      280 GGTGAGCTGCCAGATGCGGTGCGTCTTTCGCCAGCGCATGGTGACACCCACT 339
Db      335 LysLysAspIleLysAlaIleValArgAsnMetSerTyrMetGlnLysGlyThrMetThr 354
QY      340 GGCCTGGGCTGTGTTATGCCAGAACAGCTGTTTGTGAAGCATCAGTCCCGGCA 399
Db      355 GlyAlaAlaLeuLysTyrLeuIleAspAsnSerPheThrValIleSerSerGlyAlaArgPro 374
QY      400 GGGGTGCCCAAGTGTGTGTGGGTGACAGATGAGCGGCTCCAGCGACCTGTGGGCCCC 459
Db      375 GlyAlaGlnLysValGlyIleValPheThrAspGlyArgSerGlnAspTyrIleAsnAsp 394
QY      460 CCCATGACAGAGCTCAAGACCTGGGCGTCAACCGTTTCATTGTGACGACCGCGGAGGC 519
Db      395 AlaAlaLysLysAlaLysAspLeuLysPheLysMetPheAlaValGlyValGlyAsnAla 414
QY      520 AACTTCCTGGAGCTGTACGCGGCTCCCTCAGCCCT---GCCGAAAGCAGCTGCACATT 576
Db      415 ValGluAspGluLeuArgGlnIleAlaSerGluPheValAlaGlnIleTyrPheTyrThr 434
QY      577 GTGACGTGGATGACCTGACATCATTCGCCAAGAGCTGAGGGCTCCACTT 627
Db      435 AlaAspPheLysThrIleAsnGlnIleGlyLysLeuGlnLysLeuIle 451

```

RESULT 15

A37797

collagen alpha 3(VI) chain precursor - chicken

C:Species: Gallus gallus (chicken)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Aug-2004
C:Accession: A37797; A34270; A32674
R:Doliana, R.; Bonaldo, P.; Colombatti, A.
J. Cell Biol. 111, 2197-2205, 1990
A>Title: Multiple forms of chicken alpha3(VI) collagen chain generated by alternative splicing
A:Reference number: A37797; MUID:91035630; PMID:1977751
A:Accession: A37797
A:Molecule type: mRNA
A:Residues: 1-253;312-321;434-453;504-518;635-655;704-717;832-853 <DOL>
A:Cross-references: UNIPROT:P15989; UNIPARC:UPI0000173C31; UNIPARC:UPI0000173C32; UNIPARC:M24282
R:Bonaldo, P.; Russo, V.; Buccicotti, F.; Doliana, R.; Colombatti, A.
Biochemistry 29, 1245-1254, 1990
A>Title: Structural and functional features of the alpha3 chain indicate a bridging role
A:Reference number: A34270; MUID:90212613; PMID:2322559
A:Accession: A34270
A:Molecule type: mRNA
A:Residues: 224-2871 <BON>
A:Cross-references: UNIPARC:UPI0000173C38; GB:M24282
A>Note: the authors translated the codon TTC for residue 1916 as Leu and TTC for residue
J. Biol. Chem. 264, 20235-20239, 1989
A>Title: The carboxyl termini of the chicken alpha3 chain of collagen VI is a unique motif
A:Reference number: A32674; MUID:90062147; PMID:2584214
A:Accession: A32674
A:Molecule type: mRNA
A:Residues: 2151-2199;2792-3137 <BON>
A:Cross-references: UNIPARC:UPI0000173C39; UNIPARC:UPI0000173C3A; GB:M24282
C:Genetics:
A:Introns: 30/1, 236/1, 437/1; 638/1; 838/1
C:Superfamily: collagen VI

C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; extracellular
F:125/Domain: signal sequence #status predicted <SIG>
F:26-3137/Product: collagen alpha 3(VI) chain #status predicted <MAT>
F:26-2042/Domain: non-collagenous #status predicted <NMC>
F:36-202/Domain: von Willebrand factor type A repeat homology <VM01>
F:229-404/Domain: von Willebrand factor type A repeat homology <VM02>
F:442-607/Domain: von Willebrand factor type A repeat homology <VM03>
F:642-807/Domain: von Willebrand factor type A repeat homology <VM04>
F:840-1004/Domain: von Willebrand factor type A repeat homology <VM05>
F:1033-1197/Domain: von Willebrand factor type A repeat homology <VM06>
F:1237-1400/Domain: von Willebrand factor type A repeat homology <VM07>
F:1439-1604/Domain: von Willebrand factor type A repeat homology <VM08>
F:1639-1804/Domain: von Willebrand factor type A repeat homology <VM09>
F:1838-2010/Domain: von Willebrand factor type A repeat homology <VM10>
F:2043-2378/Domain: collagenous #status predicted <COL>
F:2045-2047/Region: cell attachment (R-G-D) motif
F:2153-2155/Region: cell attachment (R-G-D) motif
F:2159-2161/Region: cell attachment (R-G-D) motif
F:2379-3137/Domain: non-collagenous #status predicted <CNC>
F:2405-2577/Domain: von Willebrand factor type A repeat homology <VM11>
F:2623-2806/Domain: von Willebrand factor type A repeat homology <VM12>
F:2803-2846/Domain: platelet glycoprotein Ib-like #status predicted <GPI>
F:2954-3039/Domain: fibronectin type III repeat homology <FN3>
F:3072-3122/Domain: animal knittz-type proteinase inhibitor homology <BPI>
F:201,2084,2436,2563,2581,2683,2867,2920,3003/Binding site: cardohydrate (Asn) (covalent

Alignment Scores:

Pred. No.:	4,86e-06	Length:	3137
Score:	230.50	Matches:	76
Percent Similarity:	47.1%	Conservative:	31
Best Local Similarity:	33.5%	Mismatches:	105
Query Match:	9.7%	Indels:	15
DB:	2	Gaps:	7

US-10-699-035a-5 (1-1254) x A37797 (1-3137)

```
QY 4 CTCCTGAGACGGCG--CTCGGCTGGCCCTGAGCTTGGCGCTGGCGCGGAGC 60
DB 7 LeuProLeuAlaAlaLeuGlyLeuLeuLeuSer-----GlyPheCysSerVal 23
QY 61 GGGCGGAGCGCGGTCACAGCATCAGCCCGCA-----GGGAGCTGATG 108
DB 24 GlyAlaGlnGln-----GlnAlaAlaValArgAsnValAlaValAlaAspIleIle 40
QY 109 TTCCTGCTGAGACGCTCAGCAGCGTCTCTACACAGTTCCTCCGGGTTCCGAGATT 168
DB 41 PheLeuValAspSerSerTTPserIleGlyValGlnHisPheGlnLeuValArgGlnPhe 60
QY 169 GTGGGCGAGCTGTGGCTCCACTGCGCCCTGGGCAACCGGGCCCTGCGTGCAGTGTG 228
DB 61 LeuTyrAspValValValAlaLeuAspValGlyGlyAsnAspPheArgPheAlaLeuVal 80
QY 229 CAGTGGGCGAGTGGCGCATACCGAGTTCCTTCCTGGCCAGACAGCTGGGTGAGCT 288
DB 81 GlnPheSerGlyAsnProHisThrGlnPheGlnLeuAsnThrTyrProSerAsnGlnAsp 100
QY 289 GCCCAGGATCGGTGCGCTGCTTGGCCAGCGATGGGTGAGACACCACTGAGCTGGCG 348
DB 101 ValLeuSerHisIleAlaAsnMetProTyrMetGlyGlyGlySerIleThrGlyValGly 120
QY 349 CTGGTCTATGCGAAGAACGCTGTTGTTGTAAGCATCAGGTGCGCGG-----CCAGGG 402
DB 121 LeuGlyTyrLeuIleGlnAsnHisLeuThrValAlaGlySerArgAlaSerGlnGly 140
QY 403 GTGCCCAAGTGTGTTGGTGGTGAAGATGGCGGCTCCAGCGACCTGTGGGCCCCCCC 462
DB 141 ValProGlnValIleIleValLeuThrAspGlyGlnSerGlnAspAspValAlaLeuPro 160
QY 463 ATGAGAGAGCTCAAGGAGCTGGGCGTCAACGTTTCATTTCTGACGACCGCGGAGGAC 522
DB 161 SerSerValLeuLeuSerAlaHisValAlaAsnMetIleAlaValGlyValGlnAspAlaVal 180
QY 523 TTCCTGAGCTGTGACCGCGCTGCTCAGCCCTGCGGAGAGACCTG---CACTTTGTG 579
```

```
DB 181 GlnGlyGlnLeuLeuGlnIleAlaSerArgProPheAspThrHisLeuPheAsnLeuGln 200
QY 580 GACGTGATGACCTGCACATCATTTGTCGAAGCTGAGGGGCTCCATT---CTCGGATG 636
DB 201 AsnPheThrAlaLeuHisGlyIleValGlyAspLeuValAlaSerValArgThrSerMet 220
QY 637 CGGCCGACAGAGCTCCATGCC 657
DB 221 ThrProGlnGlnAlaGlyAla 227
```

Search completed: February 13, 2006, 13:41:49
Job time : 78.062 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:13:13 ; Search time 63.7853 Seconds
(without alignments)
2774.097 Million cell updates/sec

Title: US-10-699-035A-5

Perfect score: 2380
Sequence: 1 atgcctcccttgacgagcgct.....ccgcccagcgtgacgcgttaa 1254

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 216443 segs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL-frame+ n2p.model -DEV=xid
-DB=/abs/ABSS/REB.spool/US10699035/runat_13022006_062441_25416/app.query.fasta_1
-DB=uniprot -QFMT=fasta -SUFFIX=rup -MINMATCH=0.1 -LOOFCU=0 -LOOFEFT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USER=US10699035 @CGN 1.1 580 @runat_13022006_062441_25416 -NCPD=6 -ICPU=3
-NO MAMP -NEG SCORES=0 -WAIT -DSPBLLOC=100 -LONGLOG -REV TIMEOUT=120
-NARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Uniprot 05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2114	88.8	445	2	06PCB0_HUMAN	06PCB0 homo sapien
2	1644.5	69.1	415	2	08R225_MOUSE	08R225 mus musculu
3	1640.5	68.9	415	2	0923K3_MOUSE	0923K3 mus musculu
4	1640.5	68.9	415	2	08C007_MOUSE	08C007 mus musculu
5	1627	68.4	415	2	0642A6_RAT	0642A6 rattus norv
6	1279	53.7	281	2	0715D7_HUMAN	0715D7 homo sapien
7	1058.5	44.5	233	2	09H6D5_HUMAN	09H6D5 homo sapien
8	786	33.0	203	2	08V0V9_MOUSE	08V0V9 mus musculu
9	672.5	28.3	505	2	04SCD1_TETNG	04SCD1 tetradon n
10	421.5	18.7	3124	2	COCA1_CHICK	PI3944 gallus gall
11	421.5	17.7	2884	2	05VYK2_HUMAN	05VYK2 homo sapien
12	421.5	17.7	3063	2	05VYK1_HUMAN	05VYK1 homo sapien
13	419	17.6	3063	1	COCA1_HUMAN	05VYK1 homo sapien
14	411.5	17.3	1723	2	04SD22_TETNG	04SD22 tetradon n
15	404.5	17.0	3119	1	COCA1_MOUSE	04SD22 mus musculu
16	398.5	16.7	517	2	043853_HUMAN	043853 homo sapien

17	393	16.5	1117	2	04RXN8_TETNG	04RXN8 tetradon n
18	392	16.5	1259	2	04RP12_TETNG	04RP12 tetradon n
19	389	16.3	1297	2	04VXQ4_HUMAN	04VXQ4 homo sapien
20	389	16.3	1329	1	K1510_HUMAN	09P218 homo sapien
21	388	16.3	1284	2	04VXQ5_HUMAN	04VXQ5 homo sapien
22	388	16.3	1284	2	06P159_HUMAN	06P159 homo sapien
23	387.5	16.3	637	2	081VX1_HUMAN	081VX1 homo sapien
24	387	16.3	2944	1	CO7A1_HUMAN	097566 homo sapien
25	386.5	16.2	1253	1	097566_CANPA	097566 canis fami
26	382	16.1	1888	1	COEAL_CHICK	P32018 gallus gall
27	379.5	15.9	2936	2	07YRK8_CANPA	07YRK8 canis fami
28	374.5	15.7	2225	2	04SKX3_TETNG	04SKX3 tetradon n
29	361.5	15.2	2944	2	063870_MOUSE	063870 mus musculu
30	355.5	14.9	1472	2	0902A0_CHICK	0902A0 gallus gall
31	341.5	14.3	1797	1	COEAL_MOUSE	080X19 mus musculu
32	324	13.6	1557	2	04SH63_TETNG	04SH63 tetradon n
33	323.5	13.6	1796	1	COEAL_HUMAN	005707 homo sapien
34	322	13.5	534	2	04GOW3_HUMAN	04GOW3 homo sapien
35	317	13.3	1816	2	06NT15_BRABE	06NT15 brachydario
36	311	13.1	695	2	04NX38_9DELT	04NX38 anaeromyxob
37	310.5	13.0	1060	2	04NRA5_9DELT	04NRA5 anaeromyxob
38	291	12.2	1060	2	04NQY7_9DELT	04NQY7 anaeromyxob
39	288	12.1	828	2	04NRP2_9DELT	04NRP2 anaeromyxob
40	287	12.1	857	2	04NNU3_9DELT	04NNU3 anaeromyxob
41	285.5	12.0	2147	2	04NUB4_9DELT	04NUB4 anaeromyxob
42	283	11.9	964	2	04NU10_9DELT	04NU10 anaeromyxob
43	283	11.9	1549	2	0905K9_CHV12	0905K9 cercopithec
44	282.5	11.9	608	2	04NRA5_9DELT	04NRA5 anaeromyxob
45	282.5	11.9	2268	2	04NRA5_9DELT	04NRA5 anaeromyxob

ALIGNMENTS

RESULT 1
ID 06PCB0_HUMAN PRELIMINARY; PRT; 445 AA.
AC 06PCB0;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Von Willebrand factor A domain-related protein, isoform 1 (WARP).
GN Name=WARP; ORFNames=RP4-758J18.11-001;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RX NCBI_TaxID=9606;
RP NCBIOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hontela L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Ustin T.B., Toshynski S., Carninci P., Prange C.J.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
RA Vellalton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Feilley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Scherf A., Schen J.E., Krzywicki M.I., Skalska U., Smilans D.E.,
RA Scherf A., Schen J.E., Jones S.J.M., Maria M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
NUCLEOTIDE SEQUENCE.

RC TISSUE=Placenta;
RA Strausberg R.;
RN Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.
RP NUCLEOTIDE SEQUENCE.
RA Hall R.;
RL Submitted (May-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC059409; AAH59409.1; -; mRNA.
DR EMBL; AL391244; CA12657.1; -; Genomic DNA.
DR Ensembl; ENSG00000179403; Homo sapiens.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00041; Fm3; 2.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; Fm3; 2.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50853; Fm3; 2.
DR PROSITE; PS50234; VMPA; 1.
SQ SEQUENCE 445 AA; 46804 MW; D9FBDE9CA4DEBDF CRC64;

Alignment Scores:

Pred. No.:	2.9e-104	Length:	445
Score:	2114.00	Matches:	417
Percent Similarity:	93.7%	Conservative:	0
Best Local Similarity:	93.7%	Mismatches:	0
Query Match:	88.8%	Indels:	28
DB:	2	Gaps:	2

US-10-699-035a-5 (1-1254) x Q6PCB0_HUMAN (1-445)

QY 1 ATGCTCCCTGGAGCGCGCTCGGCGCTGAGCTTGCGGCTGCGCGCGGAGC 60
Db 1 MetLeuProTrpThrAlaLeuGlyLeuAlaLeuSerLeuAlaLeuAlaArgSer 20
QY 61 GCGCGGAGCGCGCTCCACGACATCAGCCCCCGAGGGGAGCTTCTTCTGCGAC 120
Db 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
QY 121 AGCTGAGCAGCGCTCTCTCACTACGAGTTCTCCCGGCTTCGGGAGTTTGCGGCGAGCTG 180
Db 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACTGCCCCCTGGGCAACCGGAGCGCTGCGGCAAGTCTGTGACAGTGGCAGT 240
Db 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGCGCATACCGAGTTCCTTCCGCGCACAGCTCGGCTGAGGCTGCCAGAGATGGC 300
Db 81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaGlnAspAla 100
QY 301 GTGGCGCTTCTGCGCGGCGCATGGGTGACACCCACACTGAGCTGCGCTGCTATGCG 360
Db 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisTyrGlyLeuAlaLeuValTyrAla 120
QY 361 AAGGACAGCTGTTTGTCTGAAGCATGAGTCCCGGCGAGGGGTGCCAAAGTGTGTG 420
Db 121 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuVal 140
QY 421 TGGGTGACGATGGCGGCTCCAGGACCGCTGTGGGCCCCCGCATGAGAGAGCTCAAGAGC 480
Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnLeuLysAsp 160
QY 481 CTGGGCGTACCGGTGTTCTTGTTCAGCACCGGCGCGGCAACTCTCTGAGGCTGTGAGCC 540
Db 161 LeuGlyValThrValPheHisValSerThrGlyArgGlyAspPheLeuGluLeuSerAla 180
QY 541 GCTGCTCAGCCCTTCGCGAGAGACACTGTGACTTTGTGAGCGTGATGACTGTGACATC 600
Db 181 AlaAlaSerAlaProAlaGluLysHisLeuHisPheValAspValAspLeuHisTle 200
QY 601 ATTGTTCAGAGAGTGAAGGGGCTCATTTCTC--GCATGCGGCGCGGACGACTCCATGCC 657
|||||

Db 201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
QY 658 ACGAGATCACGTCAGCGGCTTCCGCTGCGCTGCGCACCCCTGCTGACCGCAGACTCG 717
Db 221 ThrGluIleThrSerSerGlyPheArgLeuAlaTrpProProLeuLeuThrAlaAspSer 240
QY 718 GGCTCTGTGTGCTGGAGCTGTGCGCCAGCGCCGCGGGGCTGCAAGACGCCAGAG 777
Db 241 GlyTyrTyrValLeuGluLeuValProSerAlaGlnProGlyAlaAlaArgArgGlnGln 260
QY 778 CTGGCAGGAAAGCGCAGGACTGATCGTGGCGCGGCTCGACCCGAGACGAGACTACAGC 837
Db 261 LeuProGlyAsnAlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyrAsp 280
QY 838 GTGGCGTAAAGTCCGTAAGTCCAACTGCGCTCTGAGAGCCCGACATCTCGCGGTGCGC 897
Db 281 ValAlaLeuValProGluSerAsnValArgLeuLeuArgProGlnIleLeuArgValArg 300
QY 898 ACGCGG----- 903
Db 301 ThrArgProGlyGluAlaGlyProGlyAlaSerGlyProGluSerGlyAlaGlyProAla 320
QY 904 -----CCAGAGAGCGCGGCGCAGAGCGATCGTCATC 936
Db 321 ProThrGlnLeuAlaAlaLeuProAlaProGluGluAlaGlyProGluArgIleValIle 340
QY 937 TCCCAAGCGCGCGCGCAGGCTCCGCGTGAATTGGCGCCGACGCTGCGCTCAGCCGCG 996
Db 341 SerHisAlaArgProArgSerLeuArgValSerTrpAlaProAlaLeuGlySerIleAla 360
QY 997 GCGCTGCGCTACACGTCGATGTCGGGCGCTGCGGGCGGGAGCGCAGCGGCTGGAG 1056
Db 361 AlaLeuGlyTyrHisValGlnPheGlyProLeuArgGlyGlyGluAlaGlnArgValGln 380
QY 1057 GTGGCGGCGGCGCGCACTGACCACTGACGCTGAGGCGCTGCGCGCGGCGCGCTTACCTG 1116
Db 381 ValProAlaGlyArgAsnGlyThrThrLeuGlnGlyLeuAlaProGlyThrAlaTyrLeu 400
QY 1117 GTAGCGTACCGCGCTTCCGCTCGGCGCGGAGCGCGCTGCGCCAGGCGCTGCG 1176
Db 401 ValThrValThrAlaAlaPheArgSerGlyArgGluSerAlaLeuSerAlaLysAlaCys 420
QY 1177 ACGCCGACAGCG 1236
Db 421 ThrProAspGlyProArgProArgProArgProAlaProArgAlaProThrProGlyThr 440
QY 1237 GCCAGCGGTGAGCGG 1251
Db 441 AlaSerArgGluPro 445

RESULT 2
Q8R2Z5 MOUSE PRELIMINARY; PRT; 415 AA.
ID Q8R2Z5_1
AC Q8R2Z5_1
DT 01-UN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-UN-2002 (TrEMBLrel. 21, Last annotation update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Von Willebrand factor A domain-related protein (Mus musculus 11 days
DE pregnant adult female ovary and uterus cDNA, RIKEN full-length
DE enriched library, clone:5031410123 product:VON WILLEBRAND FACTOR A-
DE RELATED PROTEIN homolog).
GN Name=Vval;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Eucarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RP STRAIN=Czech II, and FVB/N;
RC TISSUE=Colon, and
RC Mammary tumor metastasized to lung. Tumor arose spontaneously;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Collins E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemkin C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gumarate P.H.,
RA Richards S., Wolley R.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
RA Fahey J., Hailton E., Keltman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buttefield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RG NIH MCC Project;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RC MEDLINE=1085660; PubMed=11217551; DOI=10.1038/35055500;
RX Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kodora K., Matsuda H.A., Ashburner M.; Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Knehl P., Lewis S., Matsuo Y., Nakaido I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kanuya M., Lee N.H.,
RA Lyons P., Marchionni L., Maehina Y., Mazzarelli J., Mombauts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilting L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.",
RT Nature 409:685-690(2001).
RN [5]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RC The PANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.",
RL Nature 420:563-573(2002).
RN [6]
RN NOCLOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RX Carninci P., Shibata Y., Hayatsu N., Sugihara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.",

Genome Res. 10:1617-1630(2000).

[7]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;

RX MEDLINE=20530913; PubMed=11076661; DOI=10.1101/gr.152600;

RA Shihata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,

RA Kono H., Akiyama J., Nishikawa T., Katsunai T., Tashiro H., Itoh M.,

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujimake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,

RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,

RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,

RT "RIXEN integrated sequence analysis (RISA) system-384-format

RT sequencing pipeline with 384 multicapillary sequencer.";

RL Genome Res. 10:11757-1771(2000).

[8]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;

RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,

RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,

RA Hayashida K., Hayatsu N., Hiramoto K., Hiroaka T., Hirozane T.,

RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,

RA Katoh H., Kawai T., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,

RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,

RA Nishi K., Nomura K., Numazaki R., Ono M., Ohashi N., Okazaki Y.,

RA Saito K., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,

RA Sasaki D., Shihata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,

RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,

RA Tamaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;

RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.

[9]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=FVB/N; TISSUE=Colon;

RA Director MGC Project;

RL Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.

DR EMBL: BC026919; AAH26919.1; -; mRNA.

DR EMBL: AK077240; BAC36703.1; -; mRNA.

DR EMBL: BC036166; AAH36166.1; -; mRNA.

DR HSSP: P18614; IMHP.

DR Ensembl: ENSMUSG0000042116; Mus musculus.

DR MGI: MGI:2179729; Wwa1.

DR GO: GO:0005615; C:extracellular space; TAS.

DR InterPro: IPR003961; FN_FIT.

DR InterPro: IPR002035; VWF_A.

DR Pfam: PF00041; Fn3_2.

DR Pfam: PF00092; VWA_1.

DR PRINTS: PR00453; VWFADOMAIN.

DR SMART: SM00060; FN3_2.

DR SMART: SM00327; VWA_1.

DR PROSITE: PS00853; FN3_2.

DR PROSITE: PS0234; VWFA_1.

SO SEQUENCE 415 AA; 44709 MW; C3887963B2E334FE CRC64;

Alignment Scores:

Pred. No.: 2,38e-79 Length: 415

Score: 1644.50 Matches: 325

Percent Similarity: 85.9% Conservative: 33

Best local Similarity: 77.9% Mismatches: 56

Query Match: 69.1% Indels: 3

DB: 2 Gaps: 1

US-10-699-035A-5 (1-1254) x QBR225_MOUSE (1-415)

OY 1 ATGCTCCCGAGCGCGCGCTTGGCGCTTGGCGCTTGGCGCGCGAGC 60

DB 1 MetLeuPheIzPrnHrAlaPheSerMetAlaLeuSerLeuAlaValaArgSer 20

OY 61 GGCAGCGAGCGCGCTTCCACGACGATCACCCCGAGGGAGACTGATGTTCTGCTGAC 120

DB 21 SerIleGluArgGlySerThrAlaSerAspProGlnGlyAspLeuLeuPheLeuAsp 40

OY 121 AGCTGACGCAAGCGTCTCTCACTACGAGATTCTCCGGGCTTGGGAGCTTGGGGCAGCTG 180

Db 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACGTGCGCCCTGGGACCGGGGCGCTGGCCAGTCTGGTGACGTGGGCACT 240
Db 61 ValAlaThrMetSerPheGlyProGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGCCATACACGAGTTCCTCCCTGGCCAGCACAGTGGGGTGAAGCTGCCAGATGG 300
Db 81 GlnProHisThrGluPheThrPheAspGlnTyrSerSerGlyGlnAlaIleGlnAspAla 100
QY 301 GTGGCGCTTCGGCCCGGACGATGGGTGACACCCACAGTGGCGCTGGCGCTGGCTGGCC 360
Db 101 IleArgValAlaProGlnThrMetGlyAspThrAsnThrGlyLeuAlaLeuAlaTyrAla 120
QY 361 AAGGAACAGCTGTTTGTCTGAAGCATGAGTCCCGGCGACGGGTGGCCAAAGTGTGTG 420
Db 121 LysGluGlnLeuPheAlaGluGlnAlaGlyAlaArgProGlyValProLysValLeuVal 140
QY 421 TGGGTGACAGATGGCGGCTTCACGACCTGTGGGCGCCCGCCATGACAGAGTCAAGAC 480
Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGluLeuLysAsp 160
QY 481 GTGGGCGTACCGGTGTTCTATTGTGACACCGCGCGGAGGAACCTTCCTGAGGTGTACGC 540
Db 161 LeuGlyValThrIlePheIleValSerThrGlyArgGlyAsnLeuLeuGluLeuLeuAla 180
QY 541 GTGCTCCACGCGCTGCGACGAGACACTGCACTTGTGAGCTGATGACCTGACATC 600
Db 181 AlaAlaSerAlaProAlaGluLysHisIlePheValAspValAspAspLeuProIle 200
QY 601 ATTGTCCAAAGAGTGAAGGAGCTTCATTCTC--GCCATGGCGCGGACGACGCTCCATGCC 657
Db 201 IleAlaArgGluLeuArgGlySerIleThrAspAlaMetGlnProGlnGlnLeuHisAla 220
QY 658 ACGGAGTACCGTCCGCGGCTTCGCGCTGGCGTGGCCACCCCGTGTGACCGGACGCTG 717
Db 221 SerGluValLeuSerSerGlyPheAspGluSerTrpProLeuLeuThrAlaAspSer 240
QY 718 GGTACTATGTGTGAGTGTGTGCCAGCGCCACCGGAGGAGCTCAAGACCGCCAGCAG 777
Db 241 GlyTyrTyrValLeuGlnLeuValProSerGlyLysLeuAlaThrThrArgGlnGln 260
QY 778 CTGCCAGGGAACGCCACGCACTGGATCTGGGCGCGCTGACCCGACACAGACTACGAC 837
Db 261 LeuProGlyAsnAlaThrSerTrpThrThrAspLeuAspProAspThrAspTyrGln 280
QY 838 GTGGCGCTGATGCTCAAGTCCAGCGCGCTGACGAGCCCGACAGTCCGCGGAGTGGCG 897
Db 281 ValSerLeuLeuProGlnSerAsnValHisLeuLeuArgProGlnHisValArgValArg 300
QY 898 ACGCGCCAGAGAGCGCGCGCGACGAGCGCATGTCATCCACGCGCGCGCGCGCAGC 957
Db 301 ThrLeuGlnGlnLunIleGlyProGlnLunArgIleValIleSerHisAlaThrProArgSer 320
QY 958 CTCCGCGTGAATTGGGCGCCAGCGCTGGGCTACGCGCGCGCTCGGATACACGTCGAC 1017
Db 321 LeuArgValSerTrpAlaProAlaLeuGlyProAspSerAlaLeuGlyTyrHisValGln 340
QY 1018 TTTCGGGCGCGTGGGCGGCGGAGGCGACGCGGTGAGGTGCGCGCGCGCGCACATCG 1077
Db 341 LeuGlyProLeuLeuGlnGlySerLeuGlnArgValGlnValProAlaGlyLunAsnSer 360
QY 1078 ACCACCTGACGAGGCTGGCGCGCGGACCGCTACCTGATGTCAGTCCGCGCGCTTC 1137
Db 361 ThrThrValGlnIleLeuThrProCysThrThrTyrLeuValThrValThrAlaAlaPhe 380
QY 1138 CGCTGGGCGCGGACGCGCTGTCTCGGACAGGCTTCGACGCGCGACGCGCGCGCGCG 1197
Db 381 ArgSerGlyArgGlnArgAlaLeuSerAlaLysAlaCysThrAlaSerGlyAlaArgThr 400
QY 1198 CGGCCACG 1244
Db 401 Arg-AlaProGlnSerMetArgProGlnAlaGlyProArgGlnPro 415

RESULT 3
Q923K3 MOUSE
ID Q923K3 MOUSE PRELIMINARY; PRT; 415 AA.
AC Q923K3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Von Willebrand factor A-related protein.
GN Name=Wval; Synonyms=4932416A1R1K, WARP;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22057805; PubMed=12062410; DOI=10.1016/S0014-5793(02)02579-6;
RA Fitzgeraid J., Ting S.T., Bateman J.P.;
RT "WARP a new member of the von Willebrand factor A-domain superfamily
of extracellular matrix proteins.";
RL FEBS Lett. 517:61-66(2002).
DR EMBL; AY030094; AAK38350.1; -, mRNA.
DR HSSP; P18614; 1MRP.
DR Ensembl; ENSMUSG00000042116; Mus musculus.
DR MGI; MGI:2179729; 4932416A1R1K.
DR MGI; MGI:2179729; Wval.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002035; vWF_A.
DR Pfam; PF00041; FN3; 2.
DR Pfam; PF00092; vWA; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00327; vWA; 1.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS50234; vWFA; 1.
SQ SEQUENCE 415 AA; 44737 MW; C3344AECB3FDE431 CRC64;

Alignment Scores:
Pred. No.: 3,88e-79 Length: 415
Score: 1640.50 Matches: 324
Percent Similarity: 85.9% Conservative: 34
Best Local Similarity: 77.7% Mismatches: 56
Query Match: 68.9% Indels: 3
DB: Gaps: 1

US-10-699-035A-5 (1-1254) x Q923K3_MOUSE (1-415)
QY 1 ATGCTCCCTGGAGCGCGCTCGGCTGCGCTTGGAGCTTGGCGCTGGCGCGGAC 60
Db 1 MetLeuPheTrpThrAlaPheSerMetAlaLeuSerLeuArgLeuAlaLeuArgSer 20
QY 61 GCGCGGAGCGCGGTTCACACAGACAGCGCCCGCGGAGGAGACATGATCTCGTCGAC 120
Db 21 SerIleGluArgGlySerThrAlaSerAspProGlnGlyAspLeuLeuPheLeuLeuAsp 40
QY 121 AGCTGACGACGCGCTCTCACTACAGATTCCTCCGAGTTCGGAGTTCGGAGGAGCTG 180
Db 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACGTGCGCCCTGGGACCGGGCGCTGCGTCCAGTCTGTGACGTGGGCACT 240
Db 61 ValAlaThrMetSerPheGlyProGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGCCATACACGAGTTCCTCCCTGGCCAGCACAGTGGGGTGAAGCTGCCAGATGG 300
Db 81 GlnProHisThrGluPheThrPheAspGlnTyrSerSerGlyGlnAlaIleGlnAspAla 100
QY 301 GTGGCGCTTCGGCCCGGACGATGGGTGACACCCACAGTGGCGCTGGCGCTGGCTGGCC 360
Db 101 IleArgValAlaProGlnThrMetGlyAspThrAsnThrGlyLeuAlaLeuAlaTyrAla 120

QY 361 AAGGACAGCTGTTGCTGACATCAGGTCGCCGCGGAGGCTGCCAAAGTGTGCTG 420
 DB 121 LysGluGlnLeuPheAlaGluGlnAlaGlyAlaArgProGlyValProLysValLeuVal 140
 QY 421 TGGGTACAGATGGCGGCTCCAGGACCCCTGGGGGCCCCCGGACGAGGCTCAAGGAC 480
 DB 141 TrpValThrAspLysGlySerSerAspProValGlyProPheGlnGlnLeuLysAsp 160
 QY 481 CTGGGGCTCAGCTGTTTCATTTGACAGCAGCGGCGGAGCAATCTTCGTGAGTGCAGCC 540
 DB 161 LeuGlyValThrLeuPheLeuValSerThrGlyArgGlyAsnLeuLeuGlnLeuLeuAla 180
 QY 541 GCTGCTCAGCCCTGCGCGGAGAACACCTGACTTTGTGACGTGATGACTGACATC 600
 DB 181 AlaAlaSerAlaProAlaGluGlyShiLeuHisPheValAlaAspLeuProLys 200
 QY 601 ATTGTCCAGAGGTGAGGGGCTCCATTCCTC--CGGATGGGCGGACGAGCTCCATGGC 657
 DB 201 IleAlaArgGlnLeuLysGlySerLleThrAspAlaMetGlnProGlnGlnLeuHisAla 220
 QY 658 ACCGAGATCAGCTCCAGCGGCTTCCGCTGAGCTGAGCCCTGCTGACCGGAGACTCG 717
 DB 221 SerGlnValLeuSerSerGlyPheAspArgLeuSerTrpProLeuLeuThrAlaAspSer 240
 QY 718 GGGTACTATGCTGAGACTGTGTCGCGGAGCCGCGGAGCGGAGCGGAGCGGAGCG 777
 DB 241 GlyTyrTyrValLeuGlnLeuValProSerGlyLysLeuAlaThrThrArgGlnGln 260
 QY 778 CTGGCGGAGAACCGGACGAGCTGATCTGGGCGGCGGCTGACCCGAGACGAGCTAGCAG 837
 DB 261 LeuProGlyAsnAlaThrSerTrpThrTrpThrAspLeuAspProAspThrAspTyrGln 280
 QY 838 GTGGCGCTAGTGTGCTGAGCTCAGCTGAGCTGAGCTGAGCGGCGGAGACTCTCGGGTGGC 897
 DB 281 ValSerLeuLeuProGlnSerAlaHisLeuLeuArgProGlnHisValAlaGlyValArg 300
 QY 898 ACCGCGGACAGAGAGCGGCGGAGCGGAGCGGAGCTGATCTCCAGCGGCGGCGGAGC 957
 DB 301 ThrLeuGlnGlnGlnAlaGlyProGlnArgGlnValIleSerHisAlaArgProArgSer 320
 QY 958 CTCCGCGTGTGTTGGGCGGCGGAGCTGAGCTGAGCGGCGGCTCGGCTACACGTCGAG 1017
 DB 321 LeuArgValSerTrpAlaProAlaLeuGlyProAspSerAlaLeuGlyTyrHisValGln 340
 QY 1018 TTGCGGCGGCTGCGGCGGAGCGGAGCGGAGCTGAGCTGAGCGGCGGCGGAGCTGCG 1077
 DB 341 LeuGlnProLeuGlnGlnGlySerLeuGlnArgValGlnValProAlaGlyGlnAsnSer 360
 QY 1078 ACCAGCCTGACGAGGCGCTGCGGCGGAGCGGAGCTGAGCTGAGCGGCGGCTTC 1137
 DB 361 ThrThrValGlnGlnLeuThrProGlyThrThrTyrLeuValThrValThrAlaAlaPhe 380
 QY 1138 CGCTCGGCGGCGGAGCGGCTGCTCGGCGGAGCGGAGCTGAGCGGCGGCGGCGGCGG 1197
 DB 381 ArgSerGlyArgGlnArgAlaLeuSerAlaValAlaCysThrAlaSerGlyAlaArgThr 400
 QY 1198 CGGCGGAGCGGCGGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGCGGCGGCGGCGG 1244
 DB 401 Arg-AlaProGlnSerMetArgProGlnAlaGlyProAlaGlnPro 415
 RESULT 4
 Q8C0Q7 MOUSE PRELIMINARY; PRT; 415 AA.
 AC 08C0Q7
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, last annotation update)
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932416A11 product:VON WILLEBRAND FACTOR A-RELATED DE PROTEIN homolog
 GN Name=vwa1; Synonym=4932416A11Rik;
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidae; Muridae; Murinae; Mus.
 OC NCBI_Taxid=10090;
 RN [1];
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RT Carninci P., Hayashizaki Y.;
 RL "High-efficiency full-length cDNA cloning."; Meth. Enzymol. 303:19-44(1999).
 [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Maehio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."; Nature 409:685-690(2001).
 RL [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA The FANTOM Consortium;
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."; Nature 420:563-573(2002).
 RL [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes."; Genome Res. 10:1617-1630(2000).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20530913; PubMed=11076661; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Katsunari T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishimi T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsubara S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., Mowman P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
 RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Bladesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherich A., Schein J.E., Jones S.J.M., Marx M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=pancreas;
 RG NIH MGC Project;
 RU Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC003543; AAH03543.2; -, mRNA.
 DR InterPro: IPR003961; FN_III.
 DR Pfam: PF00041; fn3; 2.
 DR SMART: SM00060; FN3; 1.
 DR PROSITE: PSS0853; FN3; 2.
 FT NON_TER
 SQ SEQUENCE 281 AA; 29628 MW; 350CCBA590791BD4 CRC64;

Alignment Scores:

Pred. No.: 5, 89e-60 Length: 281
 Score: 1279.00 Matches: 253
 Percent Similarity: 90.0% Conservative: 0
 Best Local Similarity: 90.0% Mismatches: 0
 Query Match: 53.7% Indels: 28
 Gaps: 2

US-10-699-035A-5 (1-1254) x Q7L5D7_HUMAN (1-281)

QY 493 GTGTTTCATTGTGACGACCGGCGAGGCAACTTCCTGAGCTGACCGCTGCTTACGCC 552
 Db 1 ValPheIleValSerThrgIyArgIyAenPheIleuGluLeuSerIleIleValGIn 20
 QY 553 CCGCGGAGACGACCTTGACCTTGTGACGCTGAGTACCTGACATCTGTCGCAAG 612
 Db 21 ProAlaGluIySerIleuHISpHeValAspValAspAspIleIleValGIn 40
 QY 613 CTGAGGGGCTCATTTCTC---GCGATCGGCGCGACACCTCCATCCACGAGATCAG 669
 Db 41 LeuArgGlySerIleuAspIleuMetArgProGInGInleuHISalThrGluIleThr 60
 QY 670 TCCAGCGGCTTCCGCTGCGCTGCGCCACCCCTGCTGACCGGACGCTGGGCTAATGTG 729
 Db 61 SerSerGlyPheArgIleuAlaTrpProIleuLeuThrAlaSerGlyTyTyVal 80
 QY 730 CTGGAGCTGGTCCGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGC 789
 Db 81 LeuGluLeuValProSerAlaGInProGlyAlaAlaArgArgInIleuProGlyAen 100
 QY 790 GCCACGAGCTGATCTGCGCGCTCTGAGGCGCCAGATCTCGGGGTGGCGCGCGG----- 903
 Db 101 AlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyAspValAlaLeuVal 120
 QY 850 CCTGAGTCCACGCTGCGCGCTCTGAGGCGCCAGATCTCGGGGTGGCGCGCGG----- 903
 Db 121 ProGluSerAsnValArgIleuAspArgProGInIleuArgValAlaGThrArgProGly 140
 QY 903 ----- 903
 Db 141 GluAlaGlyProGlyAlaSerGlyProGluSerGlyAlaGlyProAlaProThrGInIleu 160
 QY 904 -----CCAGAGAGCGCGGCGCGAGCGGATCGTCACTCTCCACGCGCGG 948
 Db 161 AlaAlaLeuProAlaProGluGInuAlaGlyProGluArgIleValIleSerHISalArg 180
 QY 949 CCGGCGAGCTCGCGCTGAGTTGGGCCCGACGCTGGGCTCAAGCCGCGGCGCTCGGCTAC 1008

Db 181 ProArgSerLeuArgValSerTrpAlaProAlaLeuGlySerAlaAlaLeuGlyTy 200
 QY 1009 CACGTGACATTCGGGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1068
 Db 201 HisValGInPheGlyProLeuArgGlyIyGlyAlaGInArgValGluValProAlaGly 220
 QY 1069 CGCAATCGACACGCTGACGAGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1128
 Db 221 ArgAsnCysThrThrLeuGInGlyLeuAlaProGlyThrAlaTyLeuValThrValThr 240
 QY 1129 GCGGCTTCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1188
 Db 241 AlaAlaPheArgSerGlyArgGluSerAlaLeuSerAlaTyAlaCysThrProAspGly 260
 QY 1189 CCGG 1248
 Db 261 ProArgProArgProArgProArgProArgProArgProArgProArgProArgProArg 280
 QY 1249 CCG 1251
 Db 281 Pro 281

RESULT 7

Q9H6J5_HUMAN
 ID Q9H6J5_HUMAN PRELIMINARY; PRT; 233 AA.
 AC Q9H6J5;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein FLN22215.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
 OC Homo.
 NC NCB1_TaxID=9606;
 RN [1]

RP NUCLEOTIDE SEQUENCE.
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
 RA Suzuki Y., Ohyaashi M., Nishi T., Shibahara T., Tanaka T.,
 RA Nakamura Y., Isgai T., Sugano S.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK025868; BAB15264.1; -, mRNA.
 DR HSP: Q96KP7; 1FNA.
 DR Ensembl: ENSG00000179403; Homo sapiens.
 DR InterPro: IPR003961; FN_III.
 DR Pfam: PF00041; fn3; 2.
 DR SMART: SM00060; FN3; 1.
 DR PROSITE: PSS0853; FN3; 2.
 SQ SEQUENCE 233 AA; 24473 MW; B2CC118FC5B7BC50 CRC64;

Alignment Scores:

Pred. No.: 2, 95e-48 Length: 233
 Score: 1058.50 Matches: 206
 Percent Similarity: 88.4% Conservative: 0
 Best Local Similarity: 88.4% Mismatches: 0
 Query Match: 44.5% Indels: 27
 Gaps: 1

US-10-699-035A-5 (1-1254) x Q9H6J5_HUMAN (1-233)

QY 634 ATGCGGCGGACAGCTCCATCCACGAGATCACGTCACGCGGCTTCGCGCTGCGCTGG 693
 Db 1 MetArgProGInGInleuHISalThrGluIleThrSerSerGlyPheArgIleuAlaTrp 20
 QY 694 CCAACCTGCTGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 753
 Db 21 ProIleuLeuThrAlaAspSerGlyTyTyValLeuGluLeuValProSerAlaGIn 40
 QY 754 CCGGGGGGCTGCAAGACCGACGAGCTGCGGAGGACCGACGAGCTGATCTGGCGCGC 813
 Db 41 ProGlyAlaAlaArgArgGInGInIleuProGlyAsnAlaThrAspTrpIleTrpAlaGly 60

QY 814 CTGAGCCGAGCAGCAGCTACGACGTGGCGCTGATGCTGACGACGCTGCTG 873
DB 61 LeuaspProaspThraspTyraPValAlaLeuValProGluSerAsnValArgLeuLeu 80
QY 874 AGGCCCCAGATCTCTGGGGGTGGCAGCGG----- 903
DB 81 ArgProGlnHleuValArgValArgThrArgProGlyGluAlaGlyProGlyAlaSerGly 100
QY 904 -----CCAGAGAG 912
DB 101 ProGluSerGlyAlaGlyProAlaProThrGlnLeuAlaLeuProAlaProGluGln 120
QY 913 GCCGGGCGCAGAGCGCATCGTATCTCCACGCCCGCGCGCAGCCTCGCGTGAAGTGG 972
DB 121 AlaGlyProGluArgGlyLeuValIleSerHisAlaArgProArgSerLeuValSerTyr 140
QY 973 GCCCGGCGCTGGGCTCAGCGCGGCGCTCGGTACACGTGAGTTCGGGCGCGTGGG 1032
DB 141 AlaProAlaLeuGlySerAlaAlaAlaLeuGlyTyrHisValGlnPheGlyProLeuArg 160
QY 1033 GCGCGGAGCGCAGCGGCTGAGTGCCTCGCGCGCGCAGCACTGACAGCGTGCAGGGC 1092
DB 161 GlyGlyGluAlaGlnArgValGlnValProAlaGlyArgAsnGlyThrThrLeuGlnGly 180
QY 1093 CTGGCGCCCGGCGCACCGCTTACTGTGACCGTGAACCGCGCTTCGCTCGGCGCGAG 1152
DB 181 LeuAlaProGlyThrAlaTyrLeuValThrValThrAlaAlaPheArgSerGlyArgGlu 200
QY 1153 AGCGGCTGTCCGCGCAAGCGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1212
DB 201 SerAlaLeuSerAlaValAlaValAlaValAlaValAlaValAlaValAlaValAlaVal 220
QY 1213 CCCCAGCG 1251
DB 221 ProArgAlaProThrProGlyThrAlaSerArgGluPro 233

RESULT 8
Q8VDV9_MOUSE PRELIMINARY; PRT; 203 AA.
ID 08VDV9_MOUSE PRELIMINARY; PRT; 203 AA.
AC 08VDV9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 4932416A11Rik protein.
GN Name=Vwal; Synonyms=4932416A11Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242608999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schlier G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang Y., Hsien F.,
RA Diatchenko L., Marusik A., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schemetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gamarallene P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heaton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Schnerch A., Schein U.E., Jones S.U.W., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RA Strausberg R.,
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC020136; AAH20136.1; mRNA.
DR Ensembl: ENSMUSG00000042116; Mus musculus.
DR MGI: MGI:2179729; 4932416A11Rik.
DR MGI: MGI:2179729; Vwal.
DR GO: GO:0005615; C:extracellular space; TAS.
DR InterPro: IPR003961; FN_III.
DR Pfam: PF00041; fn3; 2.
DR SMART: SM00060; FN3; 2.
DR PROSITE: PS00853; FN3; 2.
KW Repeat.
SQ SEQUENCE 203 AA; 22068 MW; AAD7FDEF37626898 CRC64;

Alignment Scores:
Pred. No.: 8, 52e-34 Length: 203
Score: 786.00 Matches: 152
Percent Similarity: 83.4% Conservative: 19
Best Local Similarity: 74.1% Mismatches: 32
Query Match: 33.0% Indels: 2
DB: 2 Gaps: 0

US-10-699-035A-5 (1-1254) x Q8VDV9_MOUSE (1-203)

QY 634 ATGCGCGCGCAGCAGCTCATCTCCACGAGATCACTCCAGCGGCTTCCGCTGCGCTGG 693
DB 1 MetGlnProGlnGlnLeuHisAlaSerGlyValLeuSerSerGlyPheArgLeuSerTyr 20
QY 694 CCACCCCTGTACCGCAGACACTCGGGCTACTTGTGCTGAGCTGCTGCCAGCGCCCGAG 753
DB 21 ProProLeuLeuThrAlaAspSerGlyTyrValLeuGlnLeuValProSerGlyLys 40
QY 754 CCGGGGGCTGCAAGACGCCAGCAGCTGCGAGGAAAGCCAGCAGTGTGAGCGCGCG 813
DB 41 LeuAlaThrThrAlaArgGlnGlnLeuProGlyAsnAlaThrSerTyrThrThrAsp 60
QY 814 CTGAGCCGAGCAGCAGCTACGACGTGGCGCTGATGCTGACGACGCTGCTG 873
DB 61 LeuaspProaspThraspTyraPValAlaLeuValProGluSerAsnValArgLeuLeu 80
QY 874 AGGCCCCAGATCTCTGGGGGTGGCAGCGG----- 903
DB 81 ArgProGlnHleuValArgValArgThrArgProGlyGluAlaGlyProGlyAlaSerGly 100
QY 904 -----CCAGAGAG 912
DB 101 ProGluSerGlyAlaGlyProAlaProThrGlnLeuAlaLeuProAlaProGluGln 120
QY 913 GCCGGGCGCAGAGCGCATCGTATCTCCACGCCCGCGCGCAGCCTCGCGTGAAGTGG 972
DB 121 AlaGlyProGluArgGlyLeuValIleSerHisAlaArgProArgSerLeuValSerTyr 140
QY 973 GCCCGGCGCTGGGCTCAGCGCGGCGCTCGGTACACGTGAGTTCGGGCGCGTGGG 1032
DB 141 AlaProAlaLeuGlySerAlaAlaAlaLeuGlyTyrHisValGlnPheGlyProLeuArg 160
QY 1033 GCGCGGAGCGCAGCGGCTGAGTGCCTCGCGCGCGCAGCACTGACAGCGTGCAGGGC 1092
DB 161 GlyGlyGluAlaGlnArgValGlnValProAlaGlyArgAsnGlyThrThrLeuGlnGly 180
QY 1093 CTGGCGCCCGGCGCACCGCTTACTGTGACCGTGAACCGCGCTTCGCTCGGCGCGAG 1152
DB 181 LeuAlaProGlyThrAlaTyrLeuValThrValThrAlaAlaPheArgSerGlyArgGlu 200
QY 1153 AGCGGCTGTCCGCGCAAGCGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1212
DB 201 SerAlaLeuSerAlaValAlaValAlaValAlaValAlaValAlaValAlaValAlaVal 220
QY 1213 CCCCAGCG 1251
DB 221 ProArgAlaProThrProGlyThrAlaSerArgGluPro 233

QY 934 ATCTCCACGCGCGCGCGCGCGCGCGCTCGCGTGAAGTGGAGCCGCGGCTGCAGCG 993
DB 101 IleSerHisAlaArgProArgSerLeuArgValSerTyrAlaProAlaLeuGlyProAsp 120
QY 994 GCGGCGCTGCGTACACGACGCTGCGGCGCTGCGGCGCGCGCGCGCGCGCGCGAG 1053
DB 121 SerAlaLeuGlyTyrHisValGlnLeuGlyProLeuGlnGlySerLeuGlnVal 140
QY 1054 GAGGTGCG 1113
DB 141 GluValProAlaGlyGlnIleSerThrThrValGlnGlyLeuThrProGlySerThrThr 160
QY 1114 CTGGTACCGTACGCGCGCGCTTCTGCGTGGCGCGGAGAGCGCGTGTCCGCGAGGCG 1173
DB 161 LeuValThrValThrAlaAlaPheArgSerGlyArgGlnArgAlaLeuSerAlaVal 180
QY 1174 TGCAGCGCGCAGCG 1233
DB 181 CysThrAlaSerGlyAlaArgThrArg--AlaProGlnSerMetArgProGluAlaGlyP 200
QY 1234 ACCGCGAGCGG 1244
DB 200 roArgGluPro 203

NCBI_TaxID=9031;
 [1] NUCLEOTIDE SEQUENCE.
 STRAIN=White Leghorn;
 MEDLINE=92011862; PubMed=1918137; DOI=10.1083/jcb.115.1.209;
 RA Nishida Y., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
 Yamagata M., Obata M., Kimura K.;
 "The complete primary structure of type XII collagen shows a chimeric molecule with reiterated fibronectin type III motifs, von Willebrand factor A motifs, a domain homologous to a noncollagenous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.";
 J. Cell Biol. 115:209-221(1991).
 [2] NUCLEOTIDE SEQUENCE OF 2456-3124, AND PROTEIN SEQUENCE OF 2772-2794 AND 2846-2873.
 MEDLINE=90062079; PubMed=2584192;
 RA Gordon M.K., Gerecke D.R., Dublet B., van der Reet M., Olsen B.R.;
 "Type XII collagen. A large multidomain molecule with partial homology to type IX collagen.";
 J. Biol. Chem. 264:19772-19778(1989).
 [3] NUCLEOTIDE SEQUENCE OF 2960-3076.
 MEDLINE=87317590; PubMed=3476925;
 RA Gordon M.K., Gerecke D.R., Olsen B.R.;
 "Type XII collagen: distinct extracellular matrix component discovered by cDNA cloning.";
 Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
 [4] NUCLEOTIDE SEQUENCE OF 1-1283 (ISOFORM SHORT), AND ALTERNATIVE SPLICING.
 TISSUE=embryo;
 MEDLINE=93042014; PubMed=1420368; DOI=10.1016/0167-4781(92)90145-P;
 RA Trueb J., Trueb B.;
 "The two splice variants of collagen XII share a common 5' end.";
 Biochim. Biophys. Acta 1171:97-98(1992).
 [5] ALTERNATIVE SPLICING.
 MEDLINE=95370352; PubMed=7642694; DOI=10.1083/jcb.130.4.1005;
 RA Koch M., Bohrmann B., Matthison M., Hagios C., Trueb B., Chiquet M.;
 "Large and small splice variants of collagen XII: differential expression and ligand binding.";
 J. Cell Biol. 130:1005-1014(1995).
 CC -1- FUNCTION: Type XII collagen interacts with type I collagen-containing fibrils, the COL1 domain could be associated with the surface of the fibrils, and the COL2 and NC3 domains may be localized in the pericellular matrix.
 CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of nontriple-helical sequences.
 CC -1- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Comment=The final tissue form of collagen XII may contain homotrimers of either isoform long or isoform short or any combination of isoform long and isoform short. Only isoform long is a proteoglycan. Isoform long has more restricted expression in embryonic tissue than isoform short;
 Name=Long;
 Name=Short;
 IsoId=P13944-1; Sequence=Displayed;
 Name=Short; IsoId=P13944-2; Sequence=VSP_001148;
 CC -1- TISSUE SPECIFICITY: Type XII collagen is present in tendons, ligaments, perichondrium, and perosteum, all dense connective tissues containing type I collagen.
 CC -1- DOMAIN: This sequence defines five distinct domains, two triple-helical domains (COL1 and COL2) and three nontriple-helical domains (NC1, NC2, and NC3).
 CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at each end.
 CC -1- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- PTM: O-glycosylated; glycosaminoglycan of chondroitin-sulfate type (By similarity).
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with

interrupted helices (FACIT) family.
 CC -1- SIMILARITY: Contains 18 fibronectin type-III domains.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -1- SIMILARITY: Contains 4 VWFA domains.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

 CC EMBL, D00824; BA00701.1; -; mRNA.
 CC EMBL, X61024; CAA4358.1; -; mRNA.
 CC EMBL, M17375; AAA48718.1; -; mRNA.
 CC EMBL, J05137; AAA48635.1; -; mRNA.
 CC EMBL, X67327; CAA47744.1; -; mRNA.
 CC PIR, A40020; A40020.
 CC HSSP, P56199; 1QC5.
 CC Ensembl, ENSGALG0000015908; Gallus gallus.
 CC InterPro, IPR008160; Collagen.
 CC InterPro, IPR003961; FN III.
 CC InterPro, IPR003129; Laminin_G_TSP_N.
 CC InterPro, IPR002035; VWF_A.
 CC Pfam, PF01391; Collagen_4.
 CC Pfam, PF00041; fn3; 17.
 CC Pfam, PF00092; VWFA; 4.
 CC PRINTS, PR00453; VWFADOMAIN.
 CC SMART, SM00060; FN3; 18.
 CC SMART, SM00210; TSPN; 1.
 CC SMART, SM00327; VWFA; 4.
 CC PROSITE, PS50853; FN3; 18.
 CC PROSITE, PS50234; VWFA; 4.
 CC Alternative splicing; Cell adhesion; Collagen;
 CC Direct protein sequencing; Extracellular matrix; Glycoprotein;
 CC Hydroxylation; Repeat; Signal; Structural protein.
 CC SIGNAL 1 24
 FT CHAIN 1 24
 FT DOMAIN 25 3124 Collagen alpha 1(XII) chain.
 FT DOMAIN 25 112 Fibronectin type-III 1.
 FT DOMAIN 139 311 VWFA 1.
 FT DOMAIN 332 421 Fibronectin type-III 2.
 FT DOMAIN 439 615 VWFA 2.
 FT DOMAIN 630 718 Fibronectin type-III 3.
 FT DOMAIN 721 809 Fibronectin type-III 4.
 FT DOMAIN 812 902 Fibronectin type-III 5.
 FT DOMAIN 905 993 Fibronectin type-III 6.
 FT DOMAIN 995 1083 Fibronectin type-III 7.
 FT DOMAIN 1086 1175 Fibronectin type-III 8.
 FT DOMAIN 1199 1371 VWFA 3.
 FT DOMAIN 1386 1472 Fibronectin type-III 9.
 FT DOMAIN 1474 1564 Fibronectin type-III 10.
 FT DOMAIN 1566 1654 Fibronectin type-III 11.
 FT DOMAIN 1655 1745 Fibronectin type-III 12.
 FT DOMAIN 1756 1845 Fibronectin type-III 13.
 FT DOMAIN 1847 1935 Fibronectin type-III 14.
 FT DOMAIN 1937 2026 Fibronectin type-III 15.
 FT DOMAIN 2028 2117 Fibronectin type-III 16.
 FT DOMAIN 2119 2206 Fibronectin type-III 17.
 FT DOMAIN 2210 2294 Fibronectin type-III 18.
 FT DOMAIN 2327 2500 TSP N-terminal.
 FT DOMAIN 2524 2716 Nonhelical region (NC3).
 FT REGION 2455 2750 Triple-helical region (COL2) with 1 imperfection.
 FT REGION 2751 2902
 FT REGION 2903 2945 Nonhelical region (NC2).
 FT REGION 2946 3048 Triple-helical region (COL1) with 2 imperfections.
 FT REGION 3049 3124 Nonhelical region (NC1).
 FT MOTIF 2899 2901 Cell attachment site (Potential).
 FT COMBIAS 3086 3096 Arg/Glu-rich (acidic).
 FT COMBIAS 3111 3123 Arg/Lys-rich (basic).
 FT CARBOHYD 32 32 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 797 797 O-linked (Xyl...) (Chondroitin sulfate) (Potential).

```
FT CARBOHYD 890 890 O-linked (Xyl. . .) (chondroitin sulfate)
FT FT (Potential)
FT CARBOHYD 981 981 O-linked (Xyl. . .) (chondroitin sulfate)
FT FT (Potential)
FT CARBOHYD 1006 1006 N-linked (GlcNAc. . .) (Potential)
FT CARBOHYD 1032 1032 N-linked (GlcNAc. . .) (Potential)
FT CARBOHYD 1044 1044 N-linked (GlcNAc. . .) (Potential)
FT CARBOHYD 1512 1512 N-linked (GlcNAc. . .) (Potential)
FT CARBOHYD 1767 1767 N-linked (GlcNAc. . .) (Potential)
FT CARBOHYD 2210 2210 N-linked (GlcNAc. . .) (Potential)
FT CARBOHYD 2273 2273 N-linked (GlcNAc. . .) (Potential)
FT CARBOHYD 2532 2532 N-linked (GlcNAc. . .) (Potential)
FT CARBOHYD 2683 2683 N-linked (GlcNAc. . .) (Potential)
FT VARSPLIC 25 1188 Missing (in isoform Short).
FT CONFLICT 1258 1258 /FTID=VSP_001148.
FT CONFLICT 1264 1264 T -> S (in Ref. 4).
FT CONFLICT 1264 1264 D -> E (in Ref. 4).
FT CONFLICT 2759 2759 P -> A (in Ref. 2).
FT CONFLICT 2803 2803 L -> F (in Ref. 2).
FT CONFLICT 2977 2977 V -> F (in Ref. 2).
FT CONFLICT 3075 3076 QP -> AG (in Ref. 3).
SQ SEQUENCE 3124 AA; 340582 MW; 094285AFE7F346CF CRC64;
```

Alignment Scores:

Pred. No.:	1,23e-15	Length:	3124
Score:	44.00	Matches:	123
Percent Similarity:	47.2%	Conservative:	64
Best Local Similarity:	31.1%	Mismatches:	169
Query Match:	18.7%	Indels:	40
DB:	1	Gaps:	7

US-10-699-035A-5 (1-1254) x COCA1_CHICK (1-3124)

```
QY 94 CGAGGAGACCTATGTTCTGCTGACAGCTCAGCGAGCTCTCTCACTACGAGTTCTCC 153
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 437 LysAlaAspValValPheLeuValLaepGlySerTyrSerIleGlyIleAlaAsnPheVal 456
QY 154 CGAGTTCCGAGATTGTGGGCGAGCTGTGCTCCACTGCGCCCTGGGACCGGGGCGCTG 213
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 457 LysValAlaArgAlaPheLeuGluValLeuValLysSerPheGluIleSerProArgLysVal 476
QY 214 CGTGCAGCTGCTGTCAGCTGCGGACGTCCGACATACCGGACCTCCCTCGGCCCGACAC 273
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 477 GlnIleSerLeuValIleTyrSerArgAspProIleMetGluPheSerLeuAsnArgLys 496
QY 274 AGCTCGGTGAGGCTGCTCCGAGATGCGGTGCTTCTGCTCCGACGACATGGGTGACAC 333
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 497 AsnArgValLysAspIleIleGlnAlaIleAsnThrPheProTyrArgLysSerThr 516
QY 334 CACACTGGCTGGCGCTGTCTATGCGCAAGAAACAGCTGTTGCTGACGATCAGTGGC 393
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 517 AsnThrGlyLysAlaMetThrTyrValArgGluLysValPheValThrSerLysGlySer 536
QY 394 CGGCGCAGGGGTGCGCAAGTGTGCTGCTGTCAGCGGCGCTCCGACGACCTGTG 453
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 537 ArgProAsnValProArgValMetIleLeuIleThrAspGlyLysSerSerAspAlaPhe 556
QY 454 GGGCCCCCATGACGAGCTCAAGACCTGGCGCTCAGCTGTTCTATGTCAGCACCGGC 513
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 557 LysGluProAlaIleLysLeuArgAspAlaAspValGluIlePheAlaValGlyValLys 576
QY 514 CGAGGCACTTCTGTGAGCTGTGACCGCTGCTCAGCCCTCGGACGAGACGCTGCAC 573
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 577 AspAlaValArgThrGluLeuGluAlaIleAlaSerProProAlaGluThrHisValTyr 596
QY 574 TTTTGTG---GACCTGATGATCCTGCACATCATTTGTCACAGAGCTAGAGGGCTCATCTTC 630
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 597 ThrValGluAspPheAspAlaPheGlnArgLysSerPheGluLeuThrGlnSerValCys 616
QY 631 GCGATCGCGCGCAGAGAGCTCCATGCGC----- 657
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 617 LeuArgIleGluIleGlnIleuAlaIleArgLysLysSerTyrValProAlaLysAsn 636
```

```
QY 658 -----ACGGAATACATCTCCAGCGGCTTCCGCTGGCTGGCCACCCCTGCTGACC 708
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 637 MetValPheSerAspValThrSerAspSerPheLysValSerTrpSerAlaIleGlySer 656
QY 709 GCAGACTCGGGCTACTATGCTGTGAGCTGGGCGCCAGCGCCACCGGGGCTGCAGA 768
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 657 GluGluLysSerTyrLeuIleLysTyrLysVal-----AlaIleGlyLysAspGluPhe 674
QY 769 CGCGCAGAGCTCCAGGAGAACCCACGAGCTGATGTGGGCGCGCTGACCCGAGACG 828
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 675 ILeValSerValProAlaSerSerThrSerSerValLeuThrAsnLeuLeuProGluThr 694
QY 829 GACTACGAGCTGGCGCTAGTGCCTGATGCTCAACGACGCGCTCTCTGAGGCCACGATCTCG 888
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 695 ThrTyrAlaValSerValIleAlaGlu-----Tyr 704
QY 889 CGGTGCGCAGCGCGCCAGAGAGAGCGCGGCGCGACGATGATCATCTCCACGCGCG 948
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 705 GluAspGlyAspGlyProProLeuAspGlyGluGluThrThrLeuGluValLysGlyAla 724
QY 949 CGCGCAGCTCCGCG-----GTGAGTTGGGCCCCA 978
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 725 ProArgAsnLeuArgIleThrAspGluThrThrAspSerPheIleValGlyTyrThrPro 744
QY 979 GCGCTGGCTCAGCGCGCGCGCTCGGCTACCACTGCACTTCCGCTGGCGCGCGCG 1038
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 745 AlaProGlyAsn-----ValLeuArgTyrArgLeuValTyrArgProLeuThrGly 762
QY 1039 GAGCGCGAGCGGTGAGGTGCGCGCGCGCGCACTGCAACGACCTGACGCGCGCTGGCG 1098
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 763 GluArgArgGlnValThrValSerAlaAsnGluArgSerThrThrLeuArgAsnLeuIle 782
QY 1099 CGGGACCGCGCTACCTGTGTGACCGGTGACCGCGCGCTTCCGCTGGCGCGCGAGCGCG 1158
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 783 ProAspThrArgTyrGluValSerValIleAlaGluTyrGlnSerGlyProGlyAsnAla 802
QY 1159 CTGTCCGCGCAAGCGCTGCACCGCCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1206
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 803 LeuAsnGlyTyrAlaLysThr---AspGluValArgLysAsnProArg 817
```

RESULT 11
OSVYK2 HUMAN
ID OSVYK2_HUMAN PRELIMINARY; PRT: 2884 AA.
AC OSVYK2;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Collagen, type XII, alpha 1.
GN Name=COL12A1; ORFNames=RPL-23BD15.1-003;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Tromans A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Smith M.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Corby N.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL354664; CAH71309.1; -; Genomic DNA.
DR EMBL; AL080250; CAI19897.1; -; Genomic DNA.
DR EMBL; AL096771; CAI19907.1; -; Genomic DNA.
DR EMBL; AL080250; CAH71309.1; JOINED; Genomic DNA.
DR EMBL; AL096771; CAH71309.1; JOINED; Genomic DNA.
DR EMBL; AL354664; CAI19897.1; JOINED; Genomic DNA.
DR EMBL; AL080250; CAI19907.1; JOINED; Genomic DNA.

DR EMBL; AL354664; CA119897.1; JOINED; Genomic DNA.
 DR EMBL; AL096771; CA119897.1; JOINED; Genomic DNA.
 DR EMBL; ENSG00000111799; Homo sapiens.
 DR GO; GO:0005737; Cytoplasm; IEA.
 DR GO; GO:0005578; Extracellular matrix (sensu Metazoa); IEA.
 DR GO; GO:0007155; P-cell adhesion; IEA.
 DR GO; GO:0006817; P-phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR003129; Laminin_G_TSP_N.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF00041; Fn3; 18.
 DR Pfam; PF00092; VWA; 4.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 4.
 DR PROSITE; PS00853; FN3; 18.
 DR PROSITE; PS0234; VWF_A; 4.
 DR Collagen; Extracellular matrix; Repeat; Structural protein.
 KW SEQUENCE 2884 AA; 315869 MW; 2D598F13656E45AD CRC64;

Alignment Scores:

Pred. No.:	1,92e-14	Length:	2884
Score:	421.50	Matches:	120
Percent Similarity:	46.5%	Conservative:	67
Best local Similarity:	29.9%	Mismatches:	186
Query Match:	17.7%	Indels:	29
DB:	2	Gaps:	8

US-10-699-035a-5 (1-1254) x Q5VYK2_HUMAN (1-2884)

QY 94 CGAGGGAGCTGATGTTCTTCGTCGAGACGTCAGCGCTCTCTCACTACGAGTTCTCC 153
 Db 438 LysAlaSplLeuValPheLeuValAspGlySerTyrSerIleGlyIleAlaAspPheVal 457
 QY 154 CGGGTTTCGAGAGTTTGGGGGAGCTGGTGGCTCCACTGCGCCCTGGGACCGGGGCGCTG 213
 Db 458 LysValAlaArgAlaPheLeuGluValLeuValLysSerPheGluIleSerProAlaAlaGlyVal 477
 QY 214 CGTGCAGTCTGTGTCACGTGGGACGTGGCCATACCGAGTTCCTCCCTTGGCCAGCAC 273
 Db 478 GluIleSerLeuValGlnTyrSerArgAspProIleThrGluPheThrLeuLysPhe 497
 QY 274 AGCTCGGGTGAAGCTGCCAGATGCGGTCGCTTCTCCAGCGCATGGGTGACACC 333
 Db 498 ThrLysValGluLysPheIleGluAlaIleAsnThrPheProTyrArgGlyGlySerThr 517
 QY 334 CACACTGGCTGCGGCTGCTCTATGCCAAGAACAGCTGTTGCTGAAGCATCAGGTGCC 393
 Db 518 AsnThrGlyLysAlaMetThrTyrValArgGluLysIlePheValProSerLysGlySer 537
 QY 394 CGGCGAGGGGTGCCAAAGTGTGTGTGGTGAACAGATGGCGGCTCTCCAGCGACCTGTG 453
 Db 538 ArgSerAsnValProLysValMetIleLeuIleThrAspGlyLysSerSerAspAlaPhe 557
 QY 454 GCGCCCCCATGACGAGACTCAAGAGACCTGGCGTCAACGCTTCATTCTACACACCGCC 513
 Db 558 ArgAspProAlaIleLysLeuArgAsnSerAspValGluIlePheAlaValGlyValLys 577
 QY 514 CGAGGCAACTTCTGTGAGCTGTGACGCGCTGCGCCCTCCGCGAAGACCTGCAC 573
 Db 578 AspAlaValArgSerGluLeuGluAlaIleAlaSerProAlaGluThrIleValPhe 597
 QY 574 TTGTG---GACGTGATGACCTGCACATCATTTGTCAGAGCTGAGGGCTCCATT--- 627
 Db 598 ThrValGluPhePheAspAlaPheGlnArgIleSerPheGluLeuThrGlnSerIleCys 617
 QY 628 -----CTCGCG-----ATCGGCGCGACGACG 648
 Db 618 LeuArgIleGluGluGluLeuAlaIleLysLysLysAlaTyrValProProLysAsp 637

QY 649 CTCATGTCACGAGATCACGTTCCAGCGGCTTCCGCTGAGCCACCTCTGTCGACC 708
 Db 638 LeuSerPheSerGluValThrSerTyrGlyPheLysThrAsnThrSerProAlaGlyLys 657
 QY 709 GCAGACTCGGGCTACTATGCTGTGAGCTGTGTGCCAGGCCCGCGGGGCTGTGAAGA 768
 Db 658 AsnValPheSerTyrIleIle-----ThrTyrLysGluAlaIleGly 671
 QY 769 CGCCACGACGCTG-----CCAGGAAAGCCACGAGCTGATCTGAGCGGCGCTC 816
 Db 672 AspAspGluValThrValValGluProAlaSerThrSerValValLeuSerSerLeu 691
 QY 817 GACCCGACACGAGCTACGACGCTGGCGCTGAGCTGACCTGCAACGTCGCGCTCTGAGG 876
 Db 692 LysProGluThrLeuTyrLeuValAsnValThrAlaGluTyrGluLysPheSerIle 711
 QY 877 CCCAGATCTCTGGGGTGGCGACGCGGCCAGAGAGCGCGCCACAGCGCATCTGTCATC 936
 Db 712 ProLeuAlaGlyGluGluThrThrGluGluValLysGlyAlaProArgAsnLeuLysVal 731
 QY 937 TCCACAGCCCGCGCGGCGACGCTCGCGTGAAGTGGAGCCCGACGCTGAGCTGAGCGCG 996
 Db 732 ThrAspGluThrThrAspSerPheLysIleThrTyrThrGluAlaProLys-----Arg 749
 QY 997 GCGCTGGCTACACGCTGACGTTGGGCGCTGCGGGGGGGGGGAGCGGACGCGGTGAG 1056
 Db 750 ValLeuArgTyrArgIleTyrArgProValAlaGlyGlyLysSerArgGluValThr 769
 QY 1057 GTGCCCGCGGGCGGACACTGACACGCTGACAGGGCTGGCGCGCGGACCGCTTACCTG 1116
 Db 770 ThrProAsnGlnArgArgArgThrLeuGluAsnLeuIleProAspThrTyrGlu 789
 QY 1117 GTGACCGTGAACCGCGCTTCCGCTGCGGCGGAGAGCGGCTGCGGCAAGGCGCTGC 1176
 Db 790 ValSerValIleProGluTyrPheSerGlyProGlyThrProLeuThrGlyAsnAlaIle 809
 QY 1177 AGCGCCGAGCG 1236
 Db 810 ThrGluGlu---ValArgGlyAsnProArgAspLeu---ArgValSerAspProThrThr 827
 QY 1237 GCCAGC 1242
 Db 828 SerThr 829

RESULT 12
 Q5VYK1_HUMAN
 ID Q5VYK1_HUMAN PRELIMINARY; PRT; 3063 AA.
 AC Q5VYK1;
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Collagen, type XII, alpha 1
 GN Name=COL12A1; ORFNames=RPL-238D15.1-1001;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RN NUCLEOTIDE SEQUENCE.
 RA Tromans A.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN 12
 RN NUCLEOTIDE SEQUENCE.
 RA Smith M.;
 RN Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN 13
 RN NUCLEOTIDE SEQUENCE.
 RA Corby N.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL354664; CA119897.1; -; Genomic DNA.
 DR EMBL; AL080250; CA119898.1; -; Genomic DNA.

RT chromosome 6q12-q13.
 RL Genomics 41:236-242 (1997).
 CC - FUNCTION: Type XII collagen interacts with type I collagen-
 containing fibrils, the COL1 domain could be associated with the
 surface of the fibrils, and the COL2 and NC3 domains may be
 localized in the pericellular matrix (By similarity).
 CC - SUBUNIT: trimer of identical chains each containing 190 kDa of
 nontriple-helical sequences.
 CC - ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC homotrimer of either isoform Long or isoform Short or any
 combination of isoform Long and isoform Short;
 CC Name=Long;
 CC IsoId=Q099715-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=Q099715-2; Sequence=VSP_001149;
 CC - TISSUE SPECIFICITY: Found in collagen I-containing tissues: both
 isoform Short and isoform Long appear in amnion, chorion, skeletal
 muscle, small intestine, and in cell culture of dermal
 fibroblasts, keratinocytes and endothelial cells. Only the short
 isoform is found in lung, placenta, kidney and a squamous cell
 carcinoma cell line.
 CC - PTM: The triple-helical tail is stabilized by disulfide bonds at
 each end (By similarity).
 CC - PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains (By
 similarity).
 CC - PTM: O-glycosylation of isoform Long; glycosaminoglycan of
 chondroitin-sulfate type (By similarity).
 CC - SIMILARITY: Belongs to the fibril-associated collagens with
 interrupted helices (FACIT) family.
 CC - SIMILARITY: Contains 18 fibronectin type-III domains.
 CC - SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC - SIMILARITY: Contains 4 WFPA domains.
 CC -----
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use as long as its content is in no way modified and this statement is not
 removed.
 CC -----
 CC EMBL, U73778; AAC51244.1; -; mRNA.
 DR EMBL, U73779; AAD40483.1; -; mRNA.
 DR HSSP, P18614; IMHP.
 DR Ensembl: ENSG0000011799; Homo sapiens.
 DR HGNC, HGNC:2188; COL12A1.
 DR MIM, 120320; -.
 DR GO, GO:0005595; C:collagen type XII; TAS.
 DR GO, GO:0001501; P:skeletal development; TAS.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR002035; WFPA_A.
 DR Pfam, PF01391; Collagen; 4.
 DR Pfam, PF00041; fn3; 18.
 DR Pfam, PF00092; WFPA; 4.
 DR PRINTS: PR00453; WVFADOMAIN.
 DR PROSITE, PS50853; FN3; 18.
 DR PROSITE, PS50234; WFPA; 4.
 KW Alternative splicing; Cell adhesion; Collagen;
 KW Direct protein sequencing; Extracellular matrix; Glycoprotein;
 KW Hydroxylation; Repeat; Signal; Structural protein.
 FT SIGNLU 1 24
 FT CHAIN 25 3063 Collagen alpha 1(XII) chain.
 FT DOMAIN 25 112 Fibronectin type-III 1.
 FT DOMAIN 140 316 WFPA 1.
 FT DOMAIN 333 422 Fibronectin type-III 2.
 FT DOMAIN 440 616 WFPA 2.
 FT DOMAIN 631 719 Fibronectin type-III 3.
 FT DOMAIN 722 810 Fibronectin type-III 4.
 FT DOMAIN 813 901 Fibronectin type-III 5.
 FT DOMAIN 904 993 Fibronectin type-III 6.
 FT DOMAIN 995 1083 Fibronectin type-III 7.

FT	DOMAIN	1086	1175	Fibronectin type-III 8.
FT	DOMAIN	1199	1371	WFPA 3.
FT	DOMAIN	1384	1472	Fibronectin type-III 9.
FT	DOMAIN	1563	1563	Fibronectin type-III 10.
FT	DOMAIN	1565	1562	Fibronectin type-III 11.
FT	DOMAIN	1654	1743	Fibronectin type-III 12.
FT	DOMAIN	1752	1841	Fibronectin type-III 13.
FT	DOMAIN	1843	1931	Fibronectin type-III 14.
FT	DOMAIN	1933	2022	Fibronectin type-III 15.
FT	DOMAIN	2024	2113	Fibronectin type-III 16.
FT	DOMAIN	2115	2202	Fibronectin type-III 17.
FT	DOMAIN	2206	2280	Fibronectin type-III 18.
FT	DOMAIN	2323	2496	WFPA 4.
FT	DOMAIN	2520	2712	TSP N-terminal.
FT	REGION	2451	2746	Nonhelical region (NC3).
FT	REGION	2747	2898	Triple-helical region (COL2) with 1
FT	REGION	2899	2941	imperfect.
FT	REGION	2942	3044	Nonhelical region (NC2).
FT	REGION	3045	3063	Triple-helical region (COL1) with 2
FT	REGION	3064	3064	imperfect.
FT	MOTIF	862	864	Nonhelical region (NC1).
FT	MOTIF	2779	2781	Cell attachment site (Potential).
FT	MOTIF	2895	2897	Cell attachment site (Potential).
FT	MOD_RES	2944	2944	Hydroxyproline (By similarity).
FT	MOD_RES	2947	2947	Hydroxyproline (By similarity).
FT	MOD_RES	2950	2950	Hydroxyproline (By similarity).
FT	MOD_RES	2959	2959	Hydroxyproline (By similarity).
FT	MOD_RES	2965	2965	Hydroxyproline (By similarity).
FT	MOD_RES	2968	2968	Hydroxyproline (By similarity).
FT	MOD_RES	2971	2971	Hydroxyproline (By similarity).
FT	MOD_RES	2983	2983	Hydroxyproline (By similarity).
FT	MOD_RES	3000	3000	Hydroxyproline (By similarity).
FT	MOD_RES	3003	3003	Hydroxyproline (By similarity).
FT	MOD_RES	3014	3014	Hydroxyproline (By similarity).
FT	MOD_RES	3023	3023	Hydroxyproline (By similarity).
FT	MOD_RES	3026	3026	Hydroxyproline (By similarity).
FT	MOD_RES	3029	3029	Hydroxyproline (By similarity).
FT	CARBOHYD	700	700	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	798	798	O-linked (Xyl. . .) (chondroitin sulfate)
FT	CARBOHYD	889	889	(Potential).
FT	CARBOHYD	981	981	O-linked (Xyl. . .) (chondroitin sulfate)
FT	CARBOHYD	1763	1763	(Potential).
FT	CARBOHYD	2206	2206	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	2528	2528	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	2679	2679	N-linked (GlcNAc. . .) (Potential).
FT	VARSPLIC	25	1188	Missing (in isoform Short).
FT	SEQUENCE	3063 AA;	333194 MW;	75FEA78FA8E8253 CRC64;

Alignment Scores:
 Pred. No.: 2,6e-14
 Score: 419.00
 Percent Similarity: 39.9%
 Best Local Similarity: 27.5%
 Query Match: 17.6%
 DB: 1
 Gaps: 9

US-10-699-035A-5 (1-1254) x COCAL_HUMAN (1-3063)
 94 CGAGGGACCTGATGTCCTGCGACCTGACCGAGGCTCTCATTACGAGTTCTCC 153
 Db 438 LysAlaAspIleValPheLeuValAspGlySerTyrSerIleValAlaAspPheVal 457
 Oy CGAGTTCGGAGATTGTGTGGGACGCTGTGCTCCATCCCTCGGAGACCGGGCCCTG 213
 Db 458 LysValArgAlaPheLeuValIleValValSerPheGluIleSerProAsnArgVal 477
 Oy 214 CGTGCAGTCTGTGTCGACGTCGGACGTCGACGATTCGCTTCGCGCCAGCAC 273

D	b	478	GlnIleSerLeuValGlnIleTyrSerThrArgAspProHisThrGlnPheThrLeuIleYsPhe	497
Q	y	274	AGCTCGGGGTGAGGCTGCCAGAGATCGGGTGCCTTCTGCCAGCATGGGTGACACC	333
D	b	498	ThrIysValGlnAspIleIleGlnIleIleAsnThrPheProTyrArgGlySerThr	517
Q	y	334	CACACTGGCCCTGGCCCTGGTCTATGCCAAGAACAGCTGTTGGCTGAAGCATCAGTGGC	393
D	b	518	AsnThrGlyValAlaMetThrTyrValArgGlnIlePheValProSerIysGlySer	537
Q	y	394	CGCCAGGAGGATGCCAAAGTGTGTGGGTGGAGACAGATGGCGGCTCCAGCACCTGTG	453
D	b	538	ArgSerAsnValProIysValMetIleLeuIleThrAspGlyIysSerSerAspAlaPhe	557
Q	y	454	GAGCCCCCATGACAGAGCTCAAGACCTGGCGCTACCCGTTGCTATTGTGCACCGGC	513
D	b	558	ArgAspProAlaIleTyrLeuArgAsnSerAspValGlnIlePheAlaValGlyValIys	577
Q	y	514	CGAGGCACACTCTCTGGAGCTGTACACCGCTGCTTCAGCCCTCCGGAAGACCTGCAC	573
D	b	578	AspAlaValArgSerGlnLeuGlnAlaIleAlaSerProProAlaGlnThrIleValPhe	597
Q	y	574	TTTGTG---GAGCTGGATGACCTGCACATCATTTGTCCAAGACTGAGGGCTCCATT---	627
D	b	598	ThrValGlnAspPheAspAlaPheGlnArgIleSerPheGlnLeuThrGlnSerIleCys	617
Q	y	628	-----CTGCGC-----ATGCGCGCAGCAG	648
D	b	618	LeuArgIleGlnGlnIleuLeuAlaIleIleYsIlyAlaTyrValProIleAsp	637
Q	y	649	CTCCATGCCACGGAATCATCGTCCAGCGGCTTCGGCTGGCCACCCCTG-----	702
D	b	638	LeuSerPheSerGlnValThrSerTyrGlyPheIysThrAsnTrpSerProAlaGlyIu	657
Q	y	703	-----CTGACCGCAGCATCGGGCTACATATGG	729
D	b	658	AsnValPheSerTyrHisIleThrTyrIlySerGlnAlaIleGlyAspAspGlnValThrVal	677
Q	y	730	CTGGAG-----CTGGTGGCCAGCGCCACGCG-----	756
D	b	678	ValGlnProAlaSerSerThrSerValValLeuSerSerLeuIysProGlnThrIleuYr	697
Q	y	756	-----	756
D	b	698	LeuValAsnValThrAlaGlnTyrGlnAspGlyPheSerIleProLeuAlaGlyGlnIu	717
Q	y	757	-----GGGGCTGCAGAACGCCAGCGAGCTCCAGGAGAACGCCAGCGAGC	798
D	b	718	ThrThrGlnGlnValIlySerIlyAlaProAlaGlnLeuIysValThrAspGlnThrThrAsp	737
Q	y	799	-----TGATCTGGGCG-----	810
D	b	738	SerPheIysIleThrTyrPheGlnAlaProGlyArgValLeuArgCysArgIleIleTyr	757
Q	y	810	-----	810
D	b	758	ArgProValAlaGlyGlyIleuSerArgGlnValThrThrProProAsnGlnArgArg	777
Q	y	811	-----GGCTGCAGCCGCGACCTGACGACTGACGAGCTGGCCCTAGCTGATGTCAC	861
D	b	778	ThrLeuGlnAsnLeuIleProAspThrIySTyrGlnIleSerValIleProGlnTyrPhe	797
Q	y	862	GTCGCGCTCTCTGAGGCGCCAGATCTCGCGGTGGCGACGCGGCCAGAGAGAGCGCGGCA	921
D	b	798	SerGlyProGlyIThrProLeuThrArgAsnAlaIleThrGlnGlnValArgGlyAsnPro	817
Q	y	922	GAGCGCATGTCATCTCCACGCCCGCGCGACCTCCGCTGAGTGGAGCCCGACG	981
D	b	818	ArgAspLeuArgValSerAspProThrThrSerThrMetIysLeuSerTrpSerGlyAla	837
Q	y	982	CTGGGCTCAGCGCGCGCTCGGCTCCACGTCGACGTTGGGCGCTCGCGGGCGGGAG	104
D	b	838	ProGlyIysValIlyGln-----TyrIleuValIThrTyrThrProValAlaGlyGlyGln	855

QY	1042	GGCGACGCGAGGTGAGAGGTGCCCCGCGCGGCGCACTGCACACGACGTGCAGGCGCTTGGCGCG	1101
Db	856	ThrgIngluValThrValArgGlyAspThrThrAsnThrValLeuGlnGlyLeuLysGlu	875
QY	1102	GGACACCGGCTTACCTGCGTATGACGTGACCGCGCGCTTTCGCGTCCGCGCGCGGACGAGCGCGCTG	1161
Db	876	GlyThrGlnTyrAlaLeuSerValThrAlaLeuTyrAlaSerGlyAlaGlyAspAlaLeu	895
QY	1162	TCGCGCAAGCGCTGCACG 1179	
Db	896	PheGlyGlnGlyThrThr 901	
	RESULT 14		
Q4SD22	TETNG	PRELIMINARY; PRT; 1723 AA.	
ID	Q4SD22	TETNG	
AC	Q4SD22;		
DT	13-SEP-2005	(TREMBLrel. 31, Created)	
DT	13-SEP-2005	(TREMBLrel. 31, Last sequence update)	
DT	13-SEP-2005	(TREMBLrel. 31, Last annotation update)	
DE	Chromosome 14 SCAF14645, whole genome shotgun sequence.		
DE	(Fragment).		
GN	ORFNames=GSTENG0020231001;		
OS	Tetraodon nigroviridis (Green puffer).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthopterygii; Acanthopterygii; Percomorphi; Tetraodontiformes;		
OC	Tetraodontidae; Tetraodontidae; Tetraodon.		
OX	NCBI_TaxID=99883;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Jailton O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,		
RA	Mauviel E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,		
RA	Nicod S., Jaffe D., Fisher S., Lutfalla G., Dosat C., Segreus B.,		
RA	Daetiva C., Salanoubat M., Levy M., Boudet N., Castellano S.,		
RA	Antouard C., Ubijn C., Castelli V., Kalcinka M., Vacherie B.,		
RA	Biemont C., Skalli Z., Catcolico L., Poulain J., De Bernardis V.,		
RA	Cruaud C., Duprat S., Brotier P., Coutanceau J.P., Gouzy J.,		
RA	Parra G., Landier G., Chappe C., McKernan K.J., McEwan P., Bosak S.,		
RA	Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,		
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,		
RA	Landet V., Schachter V., Queller F., Saurin W., Scarpelli C.,		
RA	Winkler P., Lander E.S., Weissbach J., Roest Crolius H.,		
RT	"Genome duplication in the teleost fish Tetraodon nigroviridis reveals		
RT	the early vertebrate proto-karyotype."		
RL	Nature 431:946-957(2004).		
RL	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RG	Genoscope, Whitehead Institute Centre for Genome Research;		
RL	Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.		
CC	-!- CAUTION: The sequence shown here is derived from an		
CC	EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; CAABE01014645; CAG01460.1; -; Genomic_DNA.		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR003962; FNIII_subd.		
DR	InterPro; IPR02035; VWF_A.		
DR	Pfam; PF00092; VWA; 3.		
DR	Pfam; PF00092; VWA; 3.		
DR	PRINTS; PR00014; FNTPEPIT.		
DR	PRINTS; PR00453; VWFADOMAIN.		
DR	SMART; SM00060; FN3; 11.		
DR	SMART; SM00327; VWA; 3.		
DR	PROSITE; PS50853; FN3; 12.		
DR	PROSITE; PS50234; VWF_A; 3.		
KW	Collagen; Extracellular matrix; Structural protein.		
FT	NON_TER 1		
FT	NON_TER 1723		
SO	SEQUENCE 1723 AA; 187496 MW; 7EB78DA80D1EB540 CRC64;		

Percent Similarity: 44.6%
 Best Local Similarity: 30.3%
 Query Match: 17.3%
 DB: 2
 Conservative: 55
 Mismatches: 183
 Indels: 29
 Gaps: 5

US-10-699-035A-5 (1-1254) x QASD22_TETNG (1-1723)

```

QY 94 CGAGGGAGCTGATGTTCTGTGTCGACAGCTGACCGAGCTCTCTGACTACGAGTTCTCC 153
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 436 GlnAlaAspValLeuLeuValAlaAspGlySerTyrSerIleGlyLeuAlaAspPheAla 455
QY 154 CGGGTTCGGAGCTTTGTGGGGAGCTGTGGCTCCACTGCCCCGCGGACCGGGGCCCTG 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 456 LysValAlaArgAlaPheLeuGluValLeuValAsnThrPheAspIleGlyProAspLysVal 475
QY 214 CGTGCAGCTGTGGTGCAGCTGGGCGAGTCCGATACAGTATCCCTTCGCGACGAC 273
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 476 GlnIleSerLeuValGlnTyrSerAlaArgAspProIleThrGluPheTyrLeuAspSerHis 495
QY 274 AGCTCGGCTGAGGCTGCCAGATGCGGTCGCTTCCGCGACGATGGGTGACACC 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 496 HisAsnLeuGlnValValThrAlaLeuArgThrPheProTyrArgGlySerThr 515
QY 334 CACACTGGCTGCGCTGCTCTATGCCAAGAAACAGCTGTTGCTGAAGCATGAGTCC 393
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 516 AsnThrGlyArgAlaMetThrTyrValArgGluThrValPheGlnAlaSerArgGlyAla 535
QY 394 CGGCGGAGGGGTGCCCAAGTGTGGTGTGTCAGACGCGGCTCCGACGACCTGTG 453
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 536 ArgAlaHisValAlaProArgValThrIleLeuIleThrAspLysSerSerAlaPhe 555
QY 454 GCGCCCCCGATGAGAGCTCAAGACACTGGCGGTCGATTCATTGTCAGACCGGC 513
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 556 GlnGluProAlaAlaAsnLeuArgAsnSerAspValGlnIlePheAlaValGlyValys 575
QY 514 CGAGGCAACTTCTGAGCTGTAGCGGCTGACGCCCTCCGCGACGACGACCTGCAC 573
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 576 AspAlaValArgSerGluLeuGluAlaIleAlaAsnAlaProAlaGluThrHisValTyr 595
QY 574 TTGTG---GACGTGATGATCCTGACATCATGTCACAGAGTGGGGCTCATTC 630
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 596 ThrValGluAspPheAspAlaPheGlnArgIleSerThrGluLeuThrGlnSerIleCys 615
QY 631 -----GCGATGCGGCGCGACGAC 648
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 616 LeuArgIleGluGlnGluLeuGlnIleIleAsnGlnArgLeuValGlnProArgAsp 635
QY 649 CTCATGACGAGATCAGCTCCAGCGGCTTCGCTGCGCTGCGCCACCCCTGTCACC 708
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 636 LeuTyrPheSerAspValGlyProArgSerPheArgAlaSerTrpGluIleAsnAlaAsn 655
QY 709 GCGAGCTCGGGCTACATGCTGCTGAGCTGTGTCCAGCGCGCCGCGGGGTGCAAGA 768
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 656 AsnValGluSerTyrLeuValGlnPheArgProThrGluGlyValAspSerHisTyrVal 675
QY 769 CGCGACAGCTGCGCAGGAAACCGCAGCACTGATCTGGCGCGCTTCACCGCAGCAGC 828
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 676 SerMetSerValProGlyAspValLeuThrAlaLeuLeuProHisLeuThrProLeuThr 695
QY 829 GACTACGACGTGGCGCTAGTGCCTGAGTCCACGTGCGCTCTGAGGCCCGCAGATCTG 888
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 696 ArgTyrGluValSerValSerAlaGlnTyrAlaTyrAlaGlyThrSerLeuProValThrGly 715
QY 889 CGGTCGCGACGCGCGCAGAGAGCGCGCGCCAGACGATCTCATCTCCACGCCGCG 948
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 716 TyrGlyThrThrAlaGluGlnArgLysSerValGlnAsnLeuValThrGluGlnSer 735
QY 949 CGCGCAGAGCTCGCGCTGAGTTGGGCCCGCAGCGCTGGGTCTCAGCCCGCGCTCGGCTAC 1008
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 736 ProGlnSerPheArgValSerTrpAlaAlaProGlyValAlaValAlaArg-----Tyr 753
QY 1009 CACGTCAAGTTCCGGCGCTGCGGGCGGGGAGCGCCACGCGGTGAGGTCCCGCGGCG 1068
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 754 ArgLeuThrTyrGlnProAlaGlyValGlyValGlnLeuGlnAlaPheThrAlaGly 773
QY 1069 CCCAACTGCAC---ACGCTGACAGGCTGCGCGCGGACCGGCTACTGTGACCGTG 1125
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 774 ProGluLeuThrMetValLeuGlnAspLeuGlnProArgThrThrTyrArgValThrVal 793
QY 1126 ACCGCGCTTCGCTCGGGCGCGAGAGCGCGCTGTCCGCAAGGCTTCAGCGCCGAC 1185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 794 ThrProGluTyrLysGlyGly-----ProGly 802
QY 1186 GCGCGCGCG 1194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 803 GlyProGln 805

RESULT 15
COCAI MOUSE
ID COCAI MOUSE STANDARD; PRT; 3119 AA.
AC 060847; P70322;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN Name=Col12a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING (ISOFORMS XIIA-1 AND
RP XIIA-1).
RC STRAIN=C57BL/6J, and Swiss Webster; TISSUE=Skin;
RX MEDLINE=96170761; PubMed=8601036;
RA Boehme K., Li Y., Oh P.S., Olsen B.R.;
RT "Primary structure of the long and short splice variants of mouse
RT collagen XII and their tissue-specific expression during embryonic
RT development.";
RL Dev. Dyn. 204:432-445(1995).
RN [2]
RP PARTIAL NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING (ISOFORMS XIIA-2
RP AND XIIA-2).
RC STRAIN=C57BL/6J; TISSUE=Skin fibroblast;
RX MEDLINE=95348349; PubMed=10419532; DOI=10.1074/jbc.274.31.22053;
RA Kanda A.M., Reichenberger E., Baur S.T., Karimbox N.Y., Taylor R.W.,
RA Olsen B.R., Nishimura I.;
RT "Structural variation of type XII collagen at its carboxyl-terminal
RT NCI domain generated by tissue-specific alternative splicing.";
RL J. Biol. Chem. 274:22053-22059(1999).
RC -1- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the pericellular matrix (by similarity).
CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences (by similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=The final tissue form of collagen XII may contain
CC homooligomers or any combination of the various isoforms;
CC Name=XIIA-1;
CC IsoId=Q60847-1; Sequence=Displayed;
CC Name=XIIA-2; Synonyms=ER#X;
CC IsoId=Q60847-2; Sequence=VSP_001151, VSP_001152;
CC Name=XIIB-1;
CC IsoId=Q60847-3; Sequence=VSP_001150;
CC Name=XIIB-2;
CC IsoId=Q60847-4; Sequence=VSP_001150, VSP_001151, VSP_001152;
CC -1- TISSUE SPECIFICITY: Highest expression in tendons, perichondrium,
CC skin, cornea, sclera, blood vessels, and peritoneum.
CC -1- DEVELOPMENTAL STAGE: The long NC3 XIIA isoforms are predominant at
CC early stages (E07 and 11); at later stages of development (E015
CC and 17) the short NC3 XIIB forms become the major forms. As the
CC short NC3 forms become the major product, the long splice variant
CC continues to be expressed in several tissues, even after birth.

```


Search completed: February 13, 2006, 13:39:21
Job time : 346.926 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:33:59 ; Search time 142.074 Seconds
(without alignments)
893.868 Million cell updates/sec

Title: US-10-699-035A-2
Perfect score: 913
Sequence: 1 RGDLMFLDSSASVSHYFES.....FVDVDDLHIVQLRGSLID 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	913	100.0	445	2 Q6PCB0 HUMAN	Q6PCB0 homo sapien
2	772	84.6	415	2 Q8R2Z5 MOUSE	Q8R2Z5 mus musculu
3	768	84.1	415	2 Q923K3 MOUSE	Q923K3 mus musculu
4	768	84.1	415	2 Q8C0Q7 MOUSE	Q8C0Q7 mus musculu
5	761	83.4	415	2 Q642A6 RAT	Q642A6 rattus norv
6	397	43.5	505	2 Q4SCD1 TETNG	Q4SCD1 tetraodon n
7	273.5	30.0	1723	2 Q4SD22 TETNG	Q4SD22 tetraodon n
8	273.5	30.0	2225	2 Q4SXEX TETNG	Q4SXEX tetraodon n
9	268.5	29.4	1259	2 Q4RP12 TETNG	Q4RP12 tetraodon n
10	265.5	29.1	1888	1 COEAL1 CHICK	P32018 gallus gall
11	265.5	29.1	3124	1 COEAL1 CHICK	P13944 gallus gall
12	265	29.0	624	1 MATN4 MOUSE	O89029 mus musculu
13	265	29.0	1557	2 Q4SH63 TETNG	Q4SH63 tetraodon n
14	263.5	28.9	2884	2 Q5VYK2 HUMAN	Q5VYK2 homo sapien
15	263.5	28.9	3063	1 COCAL1 HUMAN	O99715 homo sapien
16	263.5	28.9	3063	2 Q5VYK1 HUMAN	O5VYK1 homo sapien
17	259	28.4	622	1 MATN4 HUMAN	O95460 homo sapien
18	257.5	28.2	517	2 Q485J3 HUMAN	Q485J3 homo sapien
19	256.5	28.1	1796	1 COEAL1 HUMAN	O05707 homo sapien
20	254.5	27.9	637	2 Q81VX1 HUMAN	O81VX1 homo sapien
21	254.5	27.9	1284	2 Q4VXO5 HUMAN	O4VXO5 homo sapien
22	254.5	27.9	1284	2 Q6P159 HUMAN	Q6P159 homo sapien
23	254.5	27.9	1297	2 Q4VXO4 HUMAN	Q4VXO4 homo sapien
24	254.5	27.9	1329	1 K1510 HUMAN	O99218 homo sapien
25	251.5	27.5	451	2 Q8N2G3 HUMAN	O8N2G3 homo sapien
26	251.5	27.5	956	1 MATN2 HUMAN	O00093 homo sapien
27	251	27.5	688	2 Q4TOK3 TETNG	Q4TOK3 tetraodon n
28	250.5	27.4	955	2 Q5R9N1 PONY	Q5R9N1 pongo pygma
29	250	27.4	839	2 Q6P3N7 XENTR	Q6P3N7 xenopus tro
30	248.5	27.2	1117	2 Q4RXN8 TETNG	Q4RXN8 tetraodon n
31	248.5	27.2	1797	1 COEAL1 MOUSE	Q80X19 mus musculu

32	248	27.2	3119	1 COCAL1 MOUSE	Q60847 mus musculu
33	247.5	27.1	493	1 MATN1 CHICK	P05099 gallus gall
34	246.5	27.0	644	2 Q5NJJ1 BRARE	Q5NJJ1 brachydanto
35	246.5	27.0	821	2 Q6PYX2 BRARE	Q6PYX2 brachydanto
36	246.5	27.0	944	2 Q5NJJ5 BRARE	Q5NJJ5 brachydanto
37	245.5	26.9	534	2 Q4GOW3 HUMAN	Q4GOW3 homo sapien
38	243.5	26.7	685	2 Q5NUJ2 BRARE	Q5NUJ2 brachydanto
39	242.5	26.6	937	2 Q569V0 MOUSE	Q569V0 mus musculu
40	242.5	26.6	956	1 MATN2 MOUSE	O08746 mus musculu
41	242.5	26.6	956	2 Q8R54Z MOUSE	Q8R54Z mus musculu
42	242	26.5	490	2 Q7SYT5 XENLA	Q7SYT5 xenopus lae
43	241.5	26.5	647	2 Q4S2X7 TETNG	Q4S2X7 tetraodon n
44	241.5	26.5	1450	2 Q4RP14 TETNG	Q4RP14 tetraodon n
45	240.5	26.3	500	2 Q80VNS MOUSE	Q80VNS mus musculu

ALIGNMENTS

RESULT 1
Q6PCB0 HUMAN
ID Q6PCB0 HUMAN PRELIMINARY; PRT; 445 AA.
AC Q6PCB0;
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Von Willebrand factor A domain-related protein, isoform 1 (WARP).
GN Name=WARP; ORFNames=RP4-758J18.11-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heselt F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Schee C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.V., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield V.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherch A., Schein J.E., Jones S.J.W., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RL Hall R.;
RT Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC059409; AAHS9409.1; -; mRNA.
DR EMBL; AL391244; CA12657.1; -; Genomic DNA.
DR Ensemble; ENSG00000179403; Homo sapiens.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00092; VWA; 1.

DR PRINTS; PR00453; VMPADMAIN.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS00234; VMA; 1.
SO SEQUENCE 445 AA, 46804 MW, D9FBD83C0A4DEBDF CRC64;

Query Match 100.0%; Score 913; DB 2; Length 445;
Best Local Similarity 100.0%; Pred.No. 2.7e-72;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RGDLMFLDSSASVSHYERSRREVGQVAPLPGTGLRSLVHVSRRPTTFRPQCH 60
DB 32 RGDLMFLDSSASVSHYERSRREVGQVAPLPGTGLRSLVHVSRRPTTFRPQCH 91
OY 61 SSGEAAQDAVRASAOQMGDTHTGLALVYAKQLPAEASGAPGPKVLVMTDGGSSDPV 120
DB 92 SSGEAAQDAVRASAOQMGDTHTGLALVYAKQLPAEASGAPGPKVLVMTDGGSSDPV 151
OY 121 GPMQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLFPVDDHLHIVQELRGSLID 180
DB 152 GPMQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLFPVDDHLHIVQELRGSLID 211

RESULT 2
OBR225 MOUSE
ID OBR225 MOUSE PRELIMINARY; PRT; 415 AA.
AC OBR225;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DE Von Willebrand factor A domain-related protein (Mus musculus 11 days
DE pregnant adult female ovary and uterus cDNA, RIKEN full-length
DE enriched library, clone:5031410123 product:VON WILLEBRAND FACTOR A-
DE RELATED PROTEIN homolog).
CN Name=Wval;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II, and FVB/N;
RC TISSUE=Colon, and
RC Mammary tumor metastasized to lung. Tumor arose spontaneously;
RX MEDLINE=22388557; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,
RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marz M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RG NIH MGC Project; Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.

RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning";
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1036/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi T., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann M., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., But C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata K., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes";
RN [7]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitsuwa T., Tashiro H., Itoh M.,
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaishiji K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer";
RN [8]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hizamoto K., Hizoko T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Nunazaki R., Ono M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,

RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RN Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 [9]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=colon;
 RA Director MGC Project;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC026919; AAH26919.1; -, mRNA.
 DR EMBL; AK077240; BAC36703.1; -, mRNA.
 DR EMBL; BC036166; AAH36166.1; -, mRNA.
 DR HSSP; P18614; IMHP.
 DR Ensembl; ENSMUSG0000042116; Mus musculus.
 DR MGI; MGI:2179729; Vwal.
 DR GO; GO:0005615; Cytoplasmic space; TAS.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00041; FN3; 2.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00853; FN3; 2.
 DR PROSITE; PS50234; VWF_A; 1.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 415 AA; 44709 MW; C3887963B2E334FE CRC64;

Query Match 84.6%; Score 772; DB 2; Length 415;
 Best local Similarity 83.9%; Pred. No. 7.5e-60;
 Matches 151; Conservative 13; Mismatches 16; Indels 0; Gaps 0;
 QY 1 RQDLMLFLDSSASVSHYEFRRVREPVGQVAPLPLGTGALRASLVHVGSRPYTEPPGQH 60
 DB 32 QGDLFLDSSASVSHYEFRRVREPVGQVATWSFGPGLRSLVHVGSPHTEFTFDQY 91
 QY 61 SSGEAAQDAVRASACMGDTHTGLALVYAKEQLFAASGARRPVRYLVWTDGSSDDPV 120
 DB 92 SSGQALRDAIRVAPQRMGDTHTGLALVYAKEQLFAEAGARRPVRYLVWTDGSSDDPV 151
 QY 121 GPPMOELKDLGVTVFVSTGRGNFLLSAAASAPAEKHLHFVVDVDDLHIIVQELRGSILD 180
 DB 152 GPPMOELKDLGVTVFVSTGRGNFLLSAAASAPAEKHLHFVVDVDDLPIIARLRGSITD 211

RESULT 3
 0923K3 MOUSE
 ID 0923K3_MOUSE PRELIMINARY; PRT; 415 AA.
 AC 0923K3;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Von Willebrand factor A-related protein.
 GN Name=Wval; Synonyms=4932416A1Rik, Wasp;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22057805; PubMed=12062410; DOI=10.1016/S0014-5793(02)02579-6;
 RA Fitzgerald J., Ting S.-T., Bateman J.F.,
 RT "WASP, a new member of the von Willebrand factor A-domain superfamily
 RT of extracellular matrix proteins.";
 RL FEBS Lett. 517:61-66(2002).
 DR EMBL; AY030094; AAK8350.1; -, mRNA.
 DR HSSP; P18614; IMHP.
 DR Ensembl; ENSMUSG0000042116; Mus musculus.
 DR MGI; MGI:2179729; 4932416A1Rik.
 DR MGI; MGI:2179729; Vwal.
 DR GO; GO:0005615; Cytoplasmic space; TAS.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR002035; VWF_A.

DR Pfam; PF00041; FN3; 2.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00853; FN3; 2.
 DR PROSITE; PS50234; VWF_A; 1.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 415 AA; 44737 MW; C344AECB3FDE431 CRC64;

Query Match 84.1%; Score 768; DB 2; Length 415;
 Best local Similarity 83.3%; Pred. No. 1.7e-59;
 Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;
 QY 1 RQDLMLFLDSSASVSHYEFRRVREPVGQVAPLPLGTGALRASLVHVGSRPYTEPPGQH 60
 DB 32 QGDLFLDSSASVSHYEFRRVREPVGQVATWSFGPGLRSLVHVGSPHTEFTFDQY 91
 QY 61 SSGEAAQDAVRASACMGDTHTGLALVYAKEQLFAASGARRPVRYLVWTDGSSDDPV 120
 DB 92 SSGQALRDAIRVAPQRMGDTHTGLALVYAKEQLFAEAGARRPVRYLVWTDGSSDDPV 151
 QY 121 GPPMOELKDLGVTVFVSTGRGNFLLSAAASAPAEKHLHFVVDVDDLHIIVQELRGSILD 180
 DB 152 GPPMOELKDLGVTVFVSTGRGNFLLSAAASAPAEKHLHFVVDVDDLPIIARLRGSITD 211

RESULT 4
 08C007 MOUSE
 ID 08C007_MOUSE PRELIMINARY; PRT; 415 AA.
 AC 08C007;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male testis cDNA, Riken full-length enriched
 DE library, clone:4932416A1 product:VON WILLEBRAND FACTOR A-RELATED
 DE PROTEIN homolog.
 GN Name=Wval; Synonyms=4932416A1Rik;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai U., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa K., Hara A., Fukunishi Y., Komoto H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka N.,
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Asbunier M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsch G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima U., Mazzarelli J., Mombaur P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schonbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Borja A., Yoshida K., Hasegawa Y., Kawaji H., Kohseki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).

OY 121 GPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLD 180
 DB 152 GPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLD 211

RESULT 6
 Q4SCD1_TETNG PRELIMINARY; PRT; 505 AA.

DT 13-SEP-2005 (TREMBlrel. 31, Created)
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF14659, whole genome shotgun sequence.
 GN ORFNames=GSTENG00020556001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]

RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Mauceli E., Bounneau L., Fischer S., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Catcolico L., Poulain J., De Bernardinis V.,
 RA Biemont C., Skalli Z., Brottier P., Coutanceau J.P., Guzy J.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Landet V., Schachter V., Queller F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.,
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 CC Submitted (Feb-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL: CAAB01014659; CAC01701.1; -: Genomic DNA.
 SO SEQUENCE 505 AA; 55316 MW; C724FA399E0751FC CRC64;

Query Match 43.5%; Score 397; DB 2; Length 505;
 Best Local Similarity 44.2%; Pred. No. 1.5e-26;
 Matches 80; Conservative 35; Mismatches 64; Indels 2; Gaps 1;

OY 2 GDMLMLSSASVSHYRSRREPVGLVAPLPIGTALRASLVHVSRRPYEPFGQHS 61
 DB 37 GDVLLSSSGSVSSYSHSRMLAFSELQPSLGBDQVRVGLVGTGPRLEFGFDANA 96
 OY 62 SGEAADAVRAAGMDTHGTALVYAKEOLFPAEA--SGAPGVKVLVWTDGSSSDP 119
 DB 97 TOSLSGGSIRNKIPLRGDNITVEALKVAERVLRCGVCGAAGLPRLVWLITDGVNPD 156
 OY 120 VGPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSL 179
 DB 157 ISGPAALREBGVAIVLVSTGHSNTLVLEIYTAPEVNTLHFVDDIDMSIITDDLRLADI 216
 OY 180 D 180
 DB 217 B 217

RESULT 7
 Q4SD22_TETNG PRELIMINARY; PRT; 1723 AA.
 ID Q4SD22;
 AC Q4SD22;

DT 13-SEP-2005 (TREMBlrel. 31, Created)
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
 DE Chromosome 14 SCAF14645, whole genome shotgun sequence.
 GN ORFNames=GSTENG00020231001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]

RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Mauceli E., Bounneau L., Fischer S., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Catcolico L., Poulain J., De Bernardinis V.,
 RA Biemont C., Skalli Z., Brottier P., Coutanceau J.P., Guzy J.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Landet V., Schachter V., Queller F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.,
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 CC Submitted (Feb-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL: CAAB01014645; CAC01460.1; -: Genomic DNA.
 DR InterPro: IPR003961; FN III.
 DR InterPro: IPR003962; FNIII subd.
 DR InterPro: IPR002035; VWF_A.
 DR Pfam: PF00041; fn3; 11.
 DR Pfam: PF00092; VMA; 3.
 DR PRINTS: PR00014; ENTPELII.
 DR PRINTS: PR00453; VWFADOMAIN.
 DR SMART: SM00060; FN3; 11.
 DR SMART: SM00327; VMA; 3.
 DR PROSITE: PS50853; FN3; 12.
 DR PROSITE: PS50234; VMPA; 3.
 KW Collagen; Extracellular matrix; Structural protein.
 FT NON TER 1
 FT 1723 1723
 SO SEQUENCE 1723 AA; 187496 MW; 7EB78DA80D1E6B40 CRC64;

Query Match 30.0%; Score 273.5; DB 2; Length 1723;
 Best Local Similarity 37.4%; Pred. No. 5.2e-15;
 Matches 67; Conservative 28; Mismatches 83; Indels 1; Gaps 1;

OY 1 RGDMLFLDSSASVSHYRSRREPVGLVAPLPIGTALRASLVHVSRRPYEPFGQHS 60
 DB 436 QADVLLVDGSSYSISILANPAKRAFLVWTFDGPDKVQSLVQYSRDPHTERYLDOSH 495
 OY 61 SSGEADAVRAAGMDTHGTALVYAKEOLFPAEASGAPGVKVLVWTDGSSSDP 120
 DB 496 HNLAVYALNTRFPFRGSGSTNGRAMTVRETVFGASGARAHVRVITLLIDGSSDAF 555
 OY 121 GPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFV--DVDDLHIIVQELRGSL 178
 DB 556 QEPANILNDSVETFAVGVKDAVRSELEAIANAPAEHTHYVTEDEPDAQRISTELTOSI 614

RESULT 8
 Q4SX3_TETNG PRELIMINARY; PRT; 2225 AA.
 ID Q4SX3;
 AC Q4SX3;

AC Q4SX83;
 DT 13-SEP-2005 (TReMBLrel. 31, Created)
 DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF12445, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG00010895001;
 OS Tetradodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neocleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxId=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Maucell E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 Da Silva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 Anthonard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 Biemont C., Skalli Z., Catolico L., Poulat J., De Bernardis V.,
 Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 Lander P., Schachter V., Queller F., Saurin W., Scarpelli C.,
 Wincker P., Lander E.S., Weissbach J., Roest Crolius H.,
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype."
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; CAE01012445; CAP94689.1; -; Genomic_DNA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR003961; FN.III.
 DR InterPro; IPR001129; Laminin G TSP_N.
 DR InterPro; IPR001220; Lectin_LegB.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF00041; FN3; 8.
 DR Pfam; PF00092; VWA; 3.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 9.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 3.
 DR PROSITE; PS50853; FN3; 9.
 DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
 DR PROSITE; PS50234; VWFA; 3.
 KW Collagen; Extracellular matrix; Structural protein.
 FT NON_TER
 SQ SEQUENCE 2225 AA; 239674 MW; EC6545B60A778BD9 CRC64;
 Query Match 30.0%; Score 273.5; DB 2; Length 2225;
 Best Local Similarity 36.5%; Pred. No. 6.9e-15;
 Matches 66; Conservative 28; Mismatches 82; Indels 5; Gaps 2;
 QY 3 DLMLFLDSSASVSHYEFRRVREFGQLVAPLPGLTGALRASLVHVGSRPYTEFPFGQHS 62
 DB 8 DMVFLVDGWSIGRTNFRILVRVFLSLVAFVDDDRIRIGLAQFSGEPRLEWHLNTHTT 67
 QY 63 GEAADAYRASAQRMGDTHITGLALVYAKQLFAEASGARPGPKVLVWVTGGSSDPYGP 122
 DB 68 KEAVWEARNPYPYKGNLTLTGLALFTLENSFSPSSGSRPGIPKIGVLLTGKSDDDVIP 127
 QY 123 PMOELKDGLGVTFPIYSTGRGNFLELSAASAPAEKHLHFVDDDAHI--IYQELRGSIL 179
 DB 128 PAQRLRDAGVEFALGVKNADEGLRALIASVSEDTHTV--NVADPHLMADIVDVTLRITIC 185

QY 180 D 180
 DB 186 E 186
 RESULT 9
 Q4RP12_TETNG PRELIMINARY; PRT; 1259 AA.
 ID Q4RP12_TETNG
 AC Q4RP12;
 DT 13-SEP-2005 (TReMBLrel. 31, Created)
 DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
 DE Chromosome 10 SCAF15009, whole genome shotgun sequence.
 GN ORFNames=GSTENG00031322001;
 OS Tetradodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neocleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxId=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 Maucell E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 Da Silva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 Anthonard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 Biemont C., Skalli Z., Catolico L., Poulat J., De Bernardis V.,
 Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 Lander P., Schachter V., Queller F., Saurin W., Scarpelli C.,
 Wincker P., Lander E.S., Weissbach J., Roest Crolius H.,
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype."
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; CAE01015009; CAG09870.1; -; Genomic_DNA.
 DR InterPro; IPR001050; CAG09870.1; -; Genomic_DNA.
 DR PRINTS; PS00092; VWA; 3.
 DR SMART; SM00060; FN3; 9.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 3.
 DR PROSITE; PS50853; FN3; 9.
 DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
 KW Collagen; Extracellular matrix; Structural protein.
 FT NON_TER
 SQ SEQUENCE 1259 AA; 138606 MW; 844C01B6FE2D0E5D CRC64;
 Query Match 29.4%; Score 268.5; DB 2; Length 1259;
 Best Local Similarity 36.3%; Pred. No. 1e-14; Indels 1; Gaps 1;
 Matches 65; Conservative 29; Mismatches 84;
 QY 1 RGDLMLDSSASVSHYEFRRVREFGQLVAPLPGLTGALRASLVHVGSRPYTEFPFGQHS 60
 DB 393 QADVVLVDGWSYSLGLOAFARVAFLEVLVNSPDIQSKVGLISLYQSRDPHTPEALNTH 452
 QY 61 SGEAADAYRASAQRMGDTHITGLALVYAKQLFAEASGARPGPKVLVWVTGGSSDPV 120
 DB 453 HDINAVRAVATFPYRGGSTNTGKAMKVKXKIFASRGARQNPVRVWVLLITDGSSDSF 512
 QY 121 GPMOELKDGLGVTFPIYSTGRGNFLELSAASAPAEKHLHFVDDDAHI--IYQELRGSIL 178
 DB 513 KDAATNLNINIDVELFPAVGKDVASSELAIANPADDNVFEVDDAFORLSKELTQSI 571
 RESULT 10
 COEAL_CHICK
 ID COEAL_CHICK STANDARD; PRT; 1888 AA.
 AC P32018; Q6LBU0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Collagen alpha 1(XIV) chain precursor (Undulin).
 GN Name=COL14A1;

OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCBT_TaxID=9031;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=93185668; PubMed=8444186;
 RA Maechli C., Trub J., Kessler B., Winterhalter K.H., Trub B.;
 RT "Complete primary structure of chicken collagen XIV.";
 RL Eur. J. Biochem. 212:483-490 (1993).
 RN [2]
 RP NUCLEOTIDE SEQUENCE OF 402-1549.
 RX MEDLINE=92339443; PubMed=1339349;
 RA Trub J., Trub B.;
 RT "Type XIV collagen is a variant of undulin.";
 RL Eur. J. Biochem. 207:549-557 (1992).
 RN [3]
 RP NUCLEOTIDE SEQUENCE OF 1582-1770.
 RX MEDLINE=92037585; PubMed=1935930;
 RA Gordon M., Castagnola P., Dublet B., Linsemayer T.F.,
 RA van der Rest M., Mayne R., Olsen B.R., Penin F., Deleage G., Dublet B.,
 RT "Cloning of the cDNA for a new member of the class of fibril-associated collagens with interrupted triple helices.";
 RL Eur. J. Biochem. 201:333-338 (1991).
 RN [4]
 RP NUCLEOTIDE SEQUENCE OF 1582-1770.
 RA Apte S.S.;
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP STRUCTURE BY NMR OF 1853-1885.
 RX MEDLINE=968357967; PubMed=96945594; DOI=10.1016/S0945-053X(98)90027-0;
 RA Giry-Lozignez C., Aubert-Foucher E., Penin F., Deleage G., Dublet B.,
 RA van der Rest M.;
 RT "Identification and characterization of a heparin binding site within the NC1 domain of chicken collagen XIV.";
 RL Matrix Biol. 17:145-149 (1998).
 RN [6]
 RP STRUCTURE BY NMR OF 1852-1885.
 RX MEDLINE=99280705; PubMed=10350466; DOI=10.1021/bi9900222;
 RA Montserret R., Aubert-Foucher E., McLeish M.J., Hill J.M., Flicheux D.,
 RA Jaquinod M., van der Rest M., Deleage G., Penin F.;
 RT "Structural analysis of the heparin-binding site of the NC1 domain of collagen XIV by CD and NMR.";
 RL Biochemistry 38:6479-6488 (1999).
 CC -1- FUNCTION: An adhesive role by integrating collagen bundles. It is probably associated with the surface of interstitial collagen fibrils via COL1. The COL2 domain may then serve as a rigid arm which sticks out from the fibril and protrudes the large N-terminal globular domain into the extracellular space, where it might interact with other matrix molecules or cell surface receptors.
 CC -1- SUBUNIT: Homotrimer (Probable).
 CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -1- TISSUE SPECIFICITY: Wide tissue distribution; high presence in dense connective tissue in skeletal muscle.
 CC -1- PTM: Lysines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
 CC -1- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- PTM: May contain numerous cysteine residues involved in inter- and intramolecular disulfide bonding.
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with interrupted helices (FACIT) family.
 CC -1- SIMILARITY: Contains 8 fibronectin type-III domains.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -1- SIMILARITY: Contains 2 WFPA domains.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not

CC removed.
 DR EMBL: X70793; CAA50064.1; -; mRNA.
 DR EMBL: X70792; CAA50063.1; -; mRNA.
 DR EMBL: X66138; CAA46928.2; -; mRNA.
 DR EMBL: X65122; CAA46238.1; -; mRNA.
 DR PIR: A45974; A45974.
 DR PIR: S78476; S78476.
 DR PDB: 1B9P; NMR; A=1853-1885.
 DR PDB: 1B9Q; NMR; A=1853-1885.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR002035; WFPA_A.
 DR Pfam: PF00041; fn3; 8.
 DR Pfam: PF00092; WFPA; 2.
 DR PRINTS: PR00453; WFPA DOMAIN.
 DR PROSITE: PS50853; FN3; 8.
 DR PROSITE: PS50234; WFPA; 2.
 KW 3D-structure; Cell adhesion; Repeat; Signal; Structural protein.
 KW Glycoprotein; Hydroxylation; Potential.
 FT SIGNAL 1 28
 FT CHAIN 1 28
 FT DOMAIN 29 1888
 FT DOMAIN 158 117
 FT DOMAIN 158 330
 FT DOMAIN 352 440
 FT DOMAIN 442 532
 FT DOMAIN 534 621
 FT DOMAIN 625 711
 FT DOMAIN 742 830
 FT DOMAIN 832 921
 FT DOMAIN 925 1009
 FT DOMAIN 1042 1215
 FT DOMAIN 1239 1444
 FT REGION 1227 1468
 FT REGION 1469 1620
 FT REGION 1664 1786
 FT MOTIF 1489 1491
 FT MOTIF 1617 1619
 FT MOTIF 1617 1619
 FT COMBINS 721 733
 FT CARBOHYD 138 138
 FT CARBOHYD 1398 1398
 FT CARBOHYD 1398 1398
 FT TURN 1855 1856
 FT TURN 1859 1861
 FT HELIX 1862 1880
 FT TURN 1881 1882
 SQ SEQUENCE 1888 AA; 202668 MW; 39915BB9F46DD873 CRC64;
 Query Match 29.1%; Score 265.5; DB 1; Length 1888;
 Best Local Similarity 34.5%; Pred. No. 2.9e-14;
 Matches 61; Conservative 33; Mismatches 82; Indels 1; Gaps 1;
 QY 3 DLMFLDSSASVSHREPSRVRFEVQGLVAPLPGLGALRALSLVHVGSRPYTFPPGCHS 62
 DB 158 DIVILVDSWSIGRNFPLVRLFEFLVLSAFVSGSEKTRVGLAQYSGDPRLEHMLNAVGT 217
 QY 63 GEAADAVRASQGRGDTHTGLATVYAKEQLFAEASGAPGPVKLVWVTDCGSSDPVGP 122
 DB 218 KDVAIDAVRANLPYKKGNTLTGLATYIIENSKPAPGARPGVSKGILITDKSDDDVIP 277
 QY 123 PMQELKDLGVTVFIVSTGRGFLSLAASAPAEKGLHFV-DVDDLIITVQELRGSIT 178
 DB 278 PAKNRDAGIELFALGVKNADINELKEIASBDSITHYVAVDFNFMNSIVEGLTRTV 334
 RESULT 11
 ID COCAL CHICK STANDARD; PRT; 3124 AA.
 AC P13944; 004509;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor (Fibrochimerin).

GN Name=COL12A1;
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 CC NCBI_TaxID=9031;
 RN
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=White leghorn;
 RX MEDLINE=92011862; PubMed=1918137; DOI=10.1083/jcb.115.1.209;
 RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
 RA Nishida Y., Obara M., Kimata K.;
 RT "The complete primary structure of type XII collagen shows a chimeric
 RT molecule with reiterated fibronectin type III motifs, von Willebrand
 RT factor A motifs, a domain homologous to a noncollagenous region of
 RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp
 RT site";
 RL J. Cell Biol. 115:209-221(1991).
 RN
 RP NUCLEOTIDE SEQUENCE OF 2456-3124, AND PROTEIN SEQUENCE OF 2772-2794
 RP AND 2846-2873.
 RX MEDLINE=90062079; PubMed=2584192;
 RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
 RT "Type XII collagen. A large multidomain molecule with partial homology
 RT to type IX collagen.";
 RL J. Biol. Chem. 264:19772-19778(1989).
 RN
 RP NUCLEOTIDE SEQUENCE OF 2960-3076.
 RX MEDLINE=87317590; PubMed=3476925;
 RA Gordon M.K., Gerecke D.R., Olsen B.R.;
 RT "Type XII collagen: distinct extracellular matrix component discovered
 RT by cDNA cloning";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
 RN
 RP NUCLEOTIDE SEQUENCE OF 1-1283 (ISOFORM SHORT), AND ALTERNATIVE
 RP SPLICING.
 RC TISSUE=Embryo;
 RX MEDLINE=93042014; PubMed=1420368; DOI=10.1016/0167-4781(92)90145-P;
 RA Trueb J., Trueb B.;
 RT "The two splice variants of collagen XII share a common 5' end.";
 RL Biochim. Biophys. Acta 1171:97-98(1992).
 RN
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=95370352; PubMed=7642694; DOI=10.1083/jcb.130.4.1005;
 RA Koch M., Bohmann B., Mathison M., Hagios C., Trueb B., Chiquet M.;
 RT "Large and small splice variants of collagen XII: differential
 RT expression and ligand binding";
 RL J. Cell Biol. 130:1005-1014(1995).
 CC
 CC -1- FUNCTION: Type XII collagen interacts with type I collagen-
 CC containing fibrils, the COL1 domain could be associated with the
 CC surface of the fibrils, and the COL2 and NC3 domains may be
 CC localized in the pericellular matrix.
 CC
 CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
 CC nontriple-helical sequences.
 CC
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment: The final tissue form of collagen XII may contain
 CC homotrimers of either isoform long or isoform short or any
 CC combination of isoform long and isoform short. Only isoform long
 CC is a proteoglycan. Isoform long has more restricted expression
 CC in embryonic tissue than isoform short;
 CC Name=Long;
 CC IsoId=PI3944-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=PI3944-2; Sequence=VSP_001148;
 CC
 CC -1- TISSUE SPECIFICITY: Type XII collagen is present in tendons,
 CC ligaments, pericardium, and perosteum, all dense connective
 CC tissues containing type I collagen.
 CC
 CC -1- DOMAIN: This sequence defines five distinct domains, two triple-
 CC helical domains (COL1 and COL2) and three nontriple-helical
 CC domains (NC1, NC2, and NC3).
 CC
 CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
 CC each end.

CC
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC
 CC -1- PTM: O-glycosylated, glycosaminoglycan of chondroitin-sulfate type
 CC (By similarity).
 CC
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with
 CC interrupted helices (FACIT) family.
 CC
 CC -1- SIMILARITY: Contains 18 fibronectin type-III domains.
 CC
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC
 CC -1- SIMILARITY: Contains 4 VWFA domains.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC
 CC -----
 CC EMBL; D00824; BA00701.1; -; mRNA.
 CC EMBL; X61024; CAA43358.1; -; mRNA.
 CC EMBL; M17375; AAA48718.1; -; mRNA.
 CC EMBL; J05137; AAA48635.1; -; mRNA.
 CC EMBL; X67327; CAA47744.1; -; mRNA.
 CC F01; A40020; A40020.
 CC HSP; P56199; I0C5.
 CC Ensembl; ENSGALG00000015908; Gallus gallus.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR003961; FN III.
 CC InterPro; IPR003129; Laminin_G_TSP_N.
 CC InterPro; IPR002035; VWF_A.
 CC Pfam; PF01391; Collagen; 4.
 CC Pfam; PF00041; FN3; 17.
 CC Pfam; PF00092; VWFA; 4.
 CC PRINTS; PR00453; VWFADOMAIN.
 CC SMART; SM0060; FN3; 18.
 CC SMART; SM00210; TSPN; 1.
 CC SMART; SM00327; VWFA; 4.
 CC PROSITE; PS00653; FN3; 18.
 CC PROSITE; PS0234; VWFA; 4.
 CC Alternative splicing; Cell adhesion; Collagen;
 CC Direct protein sequencing; Extracellular matrix; Glycoprotein;
 CC Hydroxylation; Repeat; Signal; Structural protein.
 CC
 CC SIGNAL 1 24 Potential.
 CC FT CHAIN 25 3124 Collagen alpha 1(XII) chain.
 CC FT DOMAIN 25 112 Fibronectin type-III 1.
 CC FT DOMAIN 139 311 VWFA 1.
 CC FT DOMAIN 332 421 Fibronectin type-III 2.
 CC FT DOMAIN 439 615 VWFA 2.
 CC FT DOMAIN 630 718 Fibronectin type-III 3.
 CC FT DOMAIN 721 809 Fibronectin type-III 4.
 CC FT DOMAIN 812 902 Fibronectin type-III 5.
 CC FT DOMAIN 905 993 Fibronectin type-III 6.
 CC FT DOMAIN 995 1083 Fibronectin type-III 7.
 CC FT DOMAIN 1086 1175 Fibronectin type-III 8.
 CC FT DOMAIN 1199 1371 VWFA 3.
 CC FT DOMAIN 1386 1472 Fibronectin type-III 9.
 CC FT DOMAIN 1474 1564 Fibronectin type-III 10.
 CC FT DOMAIN 1566 1654 Fibronectin type-III 11.
 CC FT DOMAIN 1655 1745 Fibronectin type-III 12.
 CC FT DOMAIN 1756 1845 Fibronectin type-III 13.
 CC FT DOMAIN 1847 1935 Fibronectin type-III 14.
 CC FT DOMAIN 1937 2026 Fibronectin type-III 15.
 CC FT DOMAIN 2028 2117 Fibronectin type-III 16.
 CC FT DOMAIN 2119 2206 Fibronectin type-III 17.
 CC FT DOMAIN 2210 2294 Fibronectin type-III 18.
 CC FT DOMAIN 2327 2500 TSP N-terminal.
 CC FT DOMAIN 2524 2716 Nonhelical region (NC3).
 CC FT REGION 2455 2750 Triple-helical region (COL2) with 1
 CC FT REGION 2751 2902 imperfection.
 CC FT REGION 2903 2945 Nonhelical region (NC2).
 CC FT REGION 2946 3048 Triple-helical region (COL1) with 2
 CC FT FT imperfections.
 CC FT REGION 3049 3124 Nonhelical region (NC1).
 CC FT MOTIF 2899 2901 Cell attachment site (Potential).

FT COMBIAS 3086 3096 Asp/Glu-rich (acidic).
FT COMBIAS 3111 3123 Arg/Lys-rich (basic).
FT CARBOHYD 32 32 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 797 797 O-linked (Xyl. . .) (chondroitin sulfate)
FT CARBOHYD 890 890 O-linked (Xyl. . .) (chondroitin sulfate)
FT CARBOHYD 981 981 O-linked (Xyl. . .) (chondroitin sulfate)
FT CARBOHYD 981 981 O-linked (Xyl. . .) (chondroitin sulfate)
FT CARBOHYD 1006 1006 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 1032 1032 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 1044 1044 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 1512 1512 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 1767 1767 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 2210 2210 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 2273 2273 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 2532 2532 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 2683 2683 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 2683 2683 N-linked (GlcNAc. . .) (Potential).
FT VARSPLIC 25 1188 Missing (in isoform Short).
FT CONFLICT 1258 1258 T -> S (in Ref. 4).
FT CONFLICT 1264 1264 D -> E (in Ref. 4).
FT CONFLICT 2759 2759 P -> A (in Ref. 2).
FT CONFLICT 2803 2803 L -> F (in Ref. 2).
FT CONFLICT 2977 2977 V -> F (in Ref. 2).
FT CONFLICT 3075 3075 OP -> AG (in Ref. 3).
SQ SEQUENCE 3124 AA; 340582 MW; 094285AE7F346CF CRC64;
Query Match 29.1%; Score 265.5; DB 1; Length 3124;
Best Local Similarity 34.6%; Pred. No. 5.2e-14;
Matches 62; Conservative 33; Mismatches 83; Indels 1; Gaps 1;
QY 1 RSDLMFLDSSASVSHYEFSEVFEVGOVLPLGTGALRASLVHVSRRPYTEPFQGH 60
DB 437 KADVFLVDSGSIGIANFVKRAFLEVLKSEISPRKQVSLVGYSDPHMESELNXY 496
QY 61 SSGEBAQDAVRASAGMDETHGLALVAKEQLPASASGAPGCVKVLVWTDGSSDPV 120
DB 497 NRVKDIQAIINTEFPYSGSTNKGKMTYREKVFVTSKSGSPVPRVMTLITDGKSDAF 556
QY 121 GPPMDELKGLGTVFVSTGRGNFLELSAASAPAKKHLFV-DVDDHLIIYQELRGS 178
DB 557 KEPATLRDADVEIFAVGVKADAVTELEAIASPPATHVYVEDPAPRISFELTQSV 615
RESULT 12
MATTN4_MOUSE STANDARD; PRT; 624 AA.
ID MATTN4_MOUSE STANDARD; PRT; 624 AA.
AC 089029; 089030; 090WS3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 10-MAY-2005 (Rel. 47, Last annotation update)
DE Matrilin-4 precursor (Mat-4).
GN Name=Matn4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP NUCLEOTIDE SEQUENCE (ISOFORMS LONG AND SHORT), AND VARIANTS.
RC STRAIN=C57BL/6J, and CD-1; TISSUE=Fetal;
RX MEDLINE=98442849; PubMed=9771906; DOI=10.1016/S0014-5793(98)01111-9;
RA Wagner R., Kobbé B., Paulsson M.;
RT "Matrilin-4, a new member of the matrilin family of extracellular
matrix proteins.";
RL FEBS Lett. 436:123-127 (1998).
RN (2)
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM LONG).
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=23288257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klauser R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diachenko L.F., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan K.J., McKernan K.J., Malek J.A., Gamarate P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN (3)
RP NUCLEOTIDE SEQUENCE OF 218-257.
RX MEDLINE=99043241; PubMed=9827539; DOI=10.1016/S0014-5793(98)01293-9;
RA Wagner R., Kobbé B., Paulsson M.;
RT "Genomic organization, alternative splicing and primary structure of
human matrilin-4.";
RL FEBS Lett. 438:165-170 (1998).
CC -1- FUNCTION: Major component of the extracellular matrix of
cartilage.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
IsoId=089029-1; Sequence=Displayed;
CC -1- IsoId=089029-2; Sequence=VSP 001401;
CC -1- TISSUE SPECIFICITY: Lung, brain, sternum, kidney and heart.
CC -1- DEVELOPMENTAL STAGE: The short isoform was detected in 7 weeks old
mice but not in developing mice (19.5 dpc embryos or in 2, 8, and
21 days old animals).
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 2 WFPA domains.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC EMBL; AJ006140; CA06889.1; -; mRNA.
CC EMBL; AJ006140; CA06890.1; -; mRNA.
CC EMBL; BC036558; AA036558.1; -; mRNA.
CC EMBL; AJ010984; CA09451.1; -; Genomic DNA.
CC HSSP; P00736; IAP0.
CC Ensembl; ENSMUSG0000016995; Mus musculus.
CC WGI; WGI:1328314; Matn4.
CC GO; GO:0005615; C:extracellular space; TAS.
CC InterPro; IPR000152; Asx hydroxyl_S.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR011203; Matn_WFA.
CC InterPro; IPR02035; WFA_A.
CC Pfam; PF00008; EGF; 4.
CC Pfam; PF00092; WFA; 2.
CC PIRSF; PIRSF015217; Matn_WFA; 1.
CC PRINTS; PR00453; WFADOMAIN.
CC PROSITE; PS00010; ASX HYDROXYL; 2.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS50026; EGF_3; FALSE_NEG.
CC PROSITE; PS50234; WFA; 2.
CC KX Alternative splicing; Coiled coil; EGF-like domain; Glycoprotein;
KW Polymorphism; Repeat; Signal.
FT SIGNAL 1 21 Potential.
FT CHAIN 22 624 Matrilin-4.
FT DOMAIN 36 215 WFA 1.


```
FT DOMAIN 1655 1652 Fibronectin type-III 11.
FT DOMAIN 1654 1743 Fibronectin type-III 12.
FT DOMAIN 1752 1841 Fibronectin type-III 13.
FT DOMAIN 1843 1931 Fibronectin type-III 14.
FT DOMAIN 1933 2022 Fibronectin type-III 15.
FT DOMAIN 2024 2113 Fibronectin type-III 16.
FT DOMAIN 2115 2202 Fibronectin type-III 17.
FT DOMAIN 2206 2290 Fibronectin type-III 18.
FT DOMAIN 2323 2496 TSP N-terminal.
FT DOMAIN 2520 2712 Nonhelical region (NC3).
FT REGION 2451 2746 Triple-helical region (COL2) with 1
FT REGION 2747 2898 Triple-helical region (COL1) with 2
FT REGION 2899 2941 imperfections.
FT REGION 2942 3044 Nonhelical region (NC1).
FT REGION 3045 3063 Cell attachment site (Potential).
FT MOTIF 2779 2781 Cell attachment site (Potential).
FT MOTIF 2895 2897 Cell attachment site (Potential).
FT MOD_RES 2944 2944 Hydroxyproline (By similarity).
FT MOD_RES 2947 2947 Hydroxyproline (By similarity).
FT MOD_RES 2950 2950 Hydroxyproline (By similarity).
FT MOD_RES 2959 2959 Hydroxyproline (By similarity).
FT MOD_RES 2965 2965 Hydroxyproline (By similarity).
FT MOD_RES 2968 2968 Hydroxyproline (By similarity).
FT MOD_RES 2971 2971 Hydroxyproline (By similarity).
FT MOD_RES 2983 2983 Hydroxyproline (By similarity).
FT MOD_RES 3000 3000 Hydroxyproline (By similarity).
FT MOD_RES 3003 3003 Hydroxyproline (By similarity).
FT MOD_RES 3014 3014 Hydroxyproline (By similarity).
FT MOD_RES 3023 3023 Hydroxyproline (By similarity).
FT MOD_RES 3026 3026 Hydroxyproline (By similarity).
FT MOD_RES 3029 3029 Hydroxyproline (By similarity).
FT CARBOHYD 700 700 O-linked (Xyl...) (Potential).
FT CARBOHYD 798 798 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 889 889 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 981 981 O-linked (Xyl...) (chondroitin sulfate)
FT CARBOHYD 1763 1763 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2206 2206 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2528 2528 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2679 2679 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 25 1188 Missing (in isoform Short).
SQ SEQUENCE 3063 AA; 333194 MW; 75FBA78FA8E48293 CRC64;

Query Match 28.9%; Score 263.5; DB 1; Length 3063;
Best Local Similarity 35.8%; Pred. No. 7.7e-14;
Matches 64; Conservative 30; Mismatches 84; Indels 1; Gaps 1;
```

Search completed: February 13, 2006, 07:47:45
Job time : 144.074 secs

Claim 11; Page 76-78; 103pp; English.

The invention relates to Willebrand Factor A domain related-protein (WARP) which is a member of von Willebrand factor A (VA)-domain protein superfamily of extracellular matrix (ECM) proteins. WARP is used as a molecular marker, used for detecting a loss of ECM integrity in an animal subject, monitoring repair, regeneration or other disease processes in an animal subject and detecting a disease condition or a propensity for the development of a disease condition in an animal subject. The invention is useful for the manufacture of a medicament in the treatment of a disease condition of the ECM. The disease condition involves the cartilage, and is preferably arthritis. The invention is also used in gene therapy. The present sequence is human WARP protein

SQ Sequence 418 AA;

Alignment Scores:	
Pred. No.:	1.88e-138
Score:	2137.50
Percent Similarity:	99.8%
Best Local Similarity:	99.8%
Query Match:	89.8%
DB:	6
	length: 418
	Matches: 417
	Conservative: 0
	Mismatches: 0
	Indels: 1
	Gaps: 1

US-10-699-035A-5 (1-1254) x AAE32502 (1-418)

QY	1	ANGTCTCCCTGGACGGGCGCTCGGGCTTGAGCTTGGGGCTGGGGCGGAGC	60
Db	1	MetLeuProTThrAlaLeuGlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer	20
QY	61	GGCGGCGAGCGCGGTCCACGAGCATCAGCCCGGAGGAGGACCTGAGTTCCTGGTGC	120
Db	21	GlyAlaGlyArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp	40
QY	121	AGCTCAGCAGCGCTCTCTCACTACGATTCTCCGGGTGGGAGTTTGTGGGGCAGCTG	180
Db	41	SerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlnPheValGlyGlnLeu	60
QY	181	GTGGCTCCATGCGCCCTGGGACCGGGGGCCCTGGTGCAGTCTGGTGCACGTGGGAGT	240
Db	61	ValAlaProLeuProLeuGlyThrGlyAlaLeuAlaArgAlaSerLeuValHisValGlySer	80
QY	241	CGGCGCATACCGAGTTCCTCCCTTCGGGCGACACAGCTCGGGGTAGGCTGGCCAGGATGCG	300
Db	81	ArgProTyrThrGlnPheProPheGlyGlnHisSerSerLeuGlyAlaAlaGlnAspAla	100
QY	301	GTGCGTGTCTTCTCCGACCGCATGGGTGACACCCACACTGCGCTGGCGCTGTATAGCC	360
Db	101	ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla	120
QY	361	AAGGAACAGCTGTTTGTCTGAAGCATCAGGTCCCGGCGAGGGGTGCCAAGTCTGGTG	420
Db	121	LeuGlnGlnLeuPheAlaGlyAlaSerGlyAlaArgProGlyValProLysValLeuVal	140

Percent Similarity: 93.7%
 Best Local Similarity: 93.7%
 Query Match: 83.8%

Conservative: 0
 Mismatches: 0
 Indels: 28
 Gaps: 2

US-10-699-035a-5 (1-1254) x ADH71106 (1-445)

QY 1 ATGCTCCCTGAGCGGCGCTCGGCTGAGCTTCGAGCTTCGCGCTGCGCGGAGC 60
 DB 1 MetLeuProTrpThrAlaLeuGlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
 QY 61 GCGCGGAGGCGGCTCCACAGATACAGCCCCCGAGGGGAGCTGATGTTCTGCTGAGC 120
 DB 21 GlyAlaGlyArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
 QY 121 ACTCAGCAGCGCTCTCTCACTACAGAGTTCTCCCGGCTTCGAGAGTTTGTGGGCGAGCTG 180
 DB 41 SerSerAlaSerValSerHisTyrGlyPheSerArgValArgGlyPheValGlyGlnLeu 60
 QY 181 GTGGCTCACTGCGCCCTGAGGCGACCGGGGCTTCGCTGCGCAGTCTGAGTGCAGCTGGCACT 240
 DB 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
 QY 241 CCGGCTTAACCGAGTTCCTCCCTTCGCGCAGCAGCTCGGGTGAAGCTGCCAGAGATGCG 300
 DB 81 ArgProTyrThrGlyPheProPheGlyGlnHisSerSerGlyGlyAlaAlaGlnAspAla 100
 QY 301 GTGGCGCTTTCGCGCGGCGAGTGGGTGACACCCACACAGCTGGCGCTGCGCTGCTATGCG 360
 DB 101 ValArgAlaSerAlaGlnArgMetGlyAspPheHisTyrGlyLeuAlaLeuValIlyAla 120
 QY 361 AAGGACAGCTGTTTTCGCTGAGCATCAGGTGCGCGGCGAGGGGTGCCAAAGTCTGTG 420
 DB 121 LysGlyGlnLeuPheAlaGlyAlaSerGlyAlaArgProGlyValProLysValLeuVal 140
 QY 421 TGGGTGACAGTGGCGGCTTCAGCGACCTGTGGGCTCCCGCAGTGCAGAGCTCAAAGAC 480
 DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnLeuLysAsp 160
 QY 481 CTGGGCGTCACTGCTTCATTCGACAGCGCGCGAGGACCTTCCTGAGCTGTGAGCG 540
 DB 161 LeuGlyValThrValPheIleValSerThrGlyAspPheLeuGlnLeuSerAla 180
 QY 541 GCTGCTCAGCCCTTCGCGAGAGACCTGCACTTCGAGTGCAGTGCAGTGCAGTGC 600
 DB 181 AlaAlaSerAlaProAlaGlyLysHisPheValAspValAspAspLeuHisIle 200
 QY 601 ATTGTCAAGAGCTGAGGGCTTCATTTC---GCGATGCGGCGCGAGCACTCCATGCC 657
 DB 201 IleValGlnGlnLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
 QY 658 ACGGAGATCAGCGTTCAGCGGCTTCGCGCTGCGCGACCCCTGTCAGCCGAGCTGCG 717
 DB 221 ThrGlyIleHisSerSerGlyPheArgLeuAlaTrpProLeuLeuThrAlaAspSer 240
 QY 718 GCGTACTATGCTGTGAGCTGTGCTCCAGCGCGCAGCCGAGGCTGCAAGAGCCAGCAG 777
 DB 241 GlyTyrTyrValLeuGlnLeuValProSerAlaGlnProGlyAlaAlaArgArgGlnGln 260
 QY 778 CTGCCAGGAGCGCCAGCACTGATCTGAGCGCGCTTCGACCCGAGCACTACGAC 837
 DB 261 LeuProGlyAsnAlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyrAsp 280
 QY 838 GTGGGCTGAGTGCAGTGCAGCGGCTTCGAGGCGCCAGAGCTTCGCGGGTGGCGC 897
 DB 281 ValAlaLeuValProGlnSerHisValArgLeuLeuArgProGlnIleLeuArgValArg 300
 QY 898 ACGCGG----- 903
 DB 301 ThrArgProGlyGlyAlaGlyProGlyAlaSerGlyProGlyLysSerGlyAlaGlyProAla 320
 QY 904 -----CCAGAGAGGCGCGGCGCAGAGCGCATCGTCATC 936

DB 321 ProThrGlnLeuAlaAlaLeuProAlaProGlnGlyAlaGlyProGlnArgIleValIle 340
 QY 937 TCCCAAGCGCGCGCGAGAGCTCCGCGTGAAGTGGCCCGCAGCCGCTGAGCGCG 996
 DB 341 SerHisAlaArgProArgSerLeuArgValSerTrpAlaProAlaLeuGlySerHisAla 360
 QY 997 GCGCTCGGCTACACGTGCAAGTTCGAGGCGCTGCGGGGCGGAGGCGCAGCGGCTGAG 1056
 DB 361 AlaLeuGlyTyrHisValGlnPheGlyProLeuArgGlyGlyAlaGlnArgValGln 380
 QY 1057 GTGGCGGCGCGCGAGCACTGACCAAGCGCTGAGGCGCGCGCGCGCGCGCTACCGT 1116
 DB 381 ValProAlaGlyArgAsnCysThrThrLeuGlnGlyLeuAlaProGlyThrAlaTyrLeu 400
 QY 1117 GTACCGTACCGCGCGCTTCGCTCGGCGCGCGAGAGCGCGCTGCGCGCAAGCGCTGC 1176
 DB 401 ValThrValThrAlaAlaPheArgSerGlyArgGlnSerAlaLeuSerAlaLysAlaCys 420
 QY 1177 ACGCCGAGCG 1236
 DB 421 ThrProAspGlyProArgProArgProArgProArgProValProArgAlaProThrProGlyThr 440
 QY 1237 GCCAGCGGAGCGCG 1251
 DB 441 AlaSerArgGlnPro 445

RESULT 4
 AAE32501
 ID AAE32501 standard; protein; 415 AA.
 AC AAE32501;
 DT 24-MAR-2003 (first entry)
 XX
 DE Mouse Willebrand Factor A domain related-protein (WARP).
 XX
 KW Willebrand Factor A domain related-protein; von Willebrand Factor A; VA; extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP; gene therapy; mouse.
 XX
 OS Mus sp.
 XX
 FH Key
 FT 1..18 Location/Qualifiers
 FT /label= Signal_peptide
 FT 19..415
 FT Protein
 FT /note= "Mouse mature WARP protein"
 FT Modified-site
 FT 148
 FT /note= "O-glycosylation site"
 FT Modified-site
 FT 264
 FT /note= "N-glycosylation site"
 FT Modified-site
 FT 359
 FT /note= "N-glycosylation site"
 FT Modified-site
 FT 361
 FT /note= "O-glycosylation site"
 FT Disulfide-bond
 FT 369..393
 FT Modified-site
 FT /note= "O-glycosylation site"
 XX
 PN WO200288184-A1.
 XX
 PD 07-NOV-2002.
 XX
 PF 02-MAY-2002; 2002MO-AU000542.
 XX
 PR 02-MAY-2001; 2001AU-00004701.
 XX
 PA (MURD-) MURDOCH CHILDRENS RES INST.
 XX
 PI Bateman JF, Fitzgerald DJ;
 XX
 DR WPI; 2003-111873/10.
 DR N-PSDB; AAD50022, AAD50398.

XX kidney injury-associated molecule, KIM, polypeptides - upregulated in
 PT injured or regenerating tissues, useful to promote tissue growth and
 PT regeneration, especially to treat renal conditions.
 XX
 PS Claim 17, Page 135-137; 213pp; English.

XX The present sequence represents a kidney injury associated molecule (KIM)
 CC protein. KIM proteins can be administered therapeutically by expressing
 CC KIM encoding polynucleotides, to promote growth and/or survival of
 CC damaged tissue (e.g. renal tissue), since the KIM proteins are
 CC upregulated in injured or regenerating (especially renal) tissues. KIM
 CC fusion proteins, conjugates, antibodies and vectors can also be used
 CC therapeutically, e.g. these or the KIM proteins may be included with an
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/
 CC prophylaxis of conditions associated with dysfunction/dysregulation of
 CC KIM genes or proteins, especially renal diseases or impairments of renal
 CC function in humans (e.g. acute renal failure, acute nephritis). The
 CC polynucleotides can be used to produce antisense sequences which, when
 CC internalised into cells, can disrupt expression of a cellular KIM gene,
 CC also useful in therapy (e.g. to block the growth of tumours dependent on
 CC KIM for growth) or compositions. The proteins and polynucleotides are
 CC useful diagnostically e.g. to detect and quantify renal injury/disease
 CC (indicative of increased risk, or presence of, renal injury or impaired
 CC function), or abnormal responses to tissue injury (indicative of
 CC increased risk, or presence of, an autoimmune response or abnormal tissue
 CC growth arising from/affecting renal tissue). The proteins can also be
 CC used to locate KIM-producing cells (especially specific loci, e.g. tissue
 CC masses abnormally producing/expressing KIM such as tumours arising
 CC from/affecting renal tissue), by contacting cells with an imageable KIM-
 CC binding reagent and imaging reagent accumulation

XX
 SQ Sequence 421 AA:

Alignment Scores:

Pred. No.: 2,346-103 Length: 421
 Score: 1627.00 Matches: 324
 Percent Similarity: 84.9% Conservative: 31
 Best Local Similarity: 77.5% Mismatches: 59
 Query Match: 68.4% Indels: 4
 DB: 2 Gaps: 2

US-10-699-035A-5 (1-1254) x AAM86326 (1-421)

QY 1 ATGCTCCCTCGAGCGGCGCTGAGCTTGGGCTGGGCGGAGC 60
 DB 7 MetLeuPheThrValLeuSerMetAlaLeuSerLeuValAlaGlnSer 26
 QY 61 GGGCGGAGGCGGCTCCAGCATGAGCCCGGAGGAGCTGATGTTCTGTGAC 120
 DB 27 GlyTleGlnArgGlyProThrAlaSerAlaProGlnGlyAspLeuLeuPheLeuAsp 46
 QY 121 AGCTCAGCCAGGCTCTCTACAGAGTTTCCCGGTTGGAGATTGTTGGGCGAGCTG 180
 DB 47 SerSerAlaSerValSerHisTyrGlnPheSerAlaValArgGlnPheValGlyGlnLeu 66
 QY 181 GTGGCTCACTGCGCCGCGGACCGGGGCGCTGGTGGCGAGCTGGGAGT 240
 DB 67 ValAlaThrMetProPheGlyProGlyAlaLeuArgAlaSerLeuValHisValGlySer 86
 QY 241 CGGCGATACACCGAGTTCCCTTCGCGCAGACAGCTCGGTGAGGCTGCCAGATGCG 300
 DB 87 ArgProHisThrGlnPheThrPheAspGlnTyrSerSerGlyGlnAlaIleGlnAspAla 106
 QY 301 GTGCGGCTTCTGCGCAGCGCATGGGTGACACCCACACTGCTGGCGGTGCTGTATGCC 360
 DB 107 ValArgValAlaLabProGlnArgMetGlyAspThrAsnThrIleuAlaLeuAlaTyrAla 126
 QY 361 AAGGAACAGCTGTTGGTGAAGATCAGGTGCGCGGCGGAGGCGCAAGTGTGCTG 420
 DB 127 LysGlnGlnLeuPheAlaGlnGlnAlaGlyAlaArgLeuGlyValProLysValLeuVal 146
 QY 421 TGGGTGACAGATGCGGCTCCAGCGACCTGTGGGCGCCCGCATGACAGAGCTCAAGAC 480

DB 147 TrpValThrAspValAlaSerSerAspSerValGlyProProwetGlnGlnLeuLysAsp 166
 QY 481 CTGGGCGTCCAGCTGTTTCATTGTCAGACCGCGGAGCAACTTCTGGAGCTGTAGCC 540
 DB 167 LeuGlyValThrIlePheIleValSerThrGlyArgGlyAsnLeuLeuGlnLeuAla 186
 QY 541 GCTGCTCAGCCCTGCGCAGAGACCTGCATCTTGTGAGCGTGATGACTGCATC 600
 DB 187 AlaAlaSerAlaProAlaGlnGlySerHisLeuHisPheValAspValAspLeuProIle 206
 QY 601 ATTGTCAAGAGCTGAGGGGCTTCATTCTC--GCGATGCGGCGCGAGAGCTTCATGCC 657
 DB 207 IleAlaArgGlnLeuArgGlyAlaIleIleAspAlaMetGlnProHisGlnLeuHisAla 226
 QY 658 ACGGATACCGGCTCGCGGCTTCGCGCTGGCGTGCACCCCTGGTGCACCGAGACTCG 717
 DB 227 SerGlnIleLeuSerHisGlnPheArgLeuSerTrpProLeuLeuThrAlaAspSer 246
 QY 718 GGCTACTATGTGTGAGCTGTGTGCCAGCGCCAGCGGCGGCTGCAAGAGCCAGAG 777
 DB 247 GlyTyrTyrValLeuGlnLeuValProSerGlyLysLeuAlaIleThrArgGlnGln 266
 QY 778 CTGCGCAGGAGACCGCAGCATGATCTGGCGCGCTGCAGCCCGAGACAGATACGAC 837
 DB 267 LeuProGlyAsnAlaThrSerTrpThrTrpThrAspLeuAsnProAspThrAspTyrGlu 286
 QY 838 GTGCGCTGTGTGCTGATGTCACAGGCGCTCTGAGGCGCCAGATCTCGGAGTGGCG 897
 DB 287 ValSerLeuLeuProIleuSerHisValArgLeuLeuArgProGlnHisLeuArgValArg 306
 QY 898 ACGCGGCGCAGAGAGCGCGGCGCAGAGCGCATGTCATCTCCACCGCGCGCGCAGC 957
 DB 307 ThrLeuGlnGlnGlnAlaGlyProGlnArgGlnIleValIleSerHisThrArgProArgSer 326
 QY 958 CTCCGCTGATGTTGGCGCCCGCAGCGCTGAGCGCTGACCGCGGCTGCTACACGTCGAG 1017
 DB 327 LeuArgValSerTrpAlaProAlaLeuGlyProAspSerThrLeuGlyTyrLeuValGln 346
 QY 1018 TTGCGGCGCTGCGGCGGCGGAGCGAGCGGCTGAGGCTGCGCGGCGCGCACTCG 1077
 DB 347 LeuGlyProLeuGlnIleGlySerLeuGlnHisValGlnValProAlaGlyIleAsnSer 366
 QY 1078 ACCACGCTGCGAGGCGCTGCGCGCGCAGCGCTTACCTGATGACCGCGCTTC 1137
 DB 367 ThrThrIleGlnGlyLeuThrProGlySerThrThrTyrLeuValThrAlaIlePhe 386
 QY 1138 CGCTCGGCGCGCAGAGCGCGCTGTGCGCCAGGCTGACCGCGCAGCGCGCGCG 1197
 DB 387 ArgSerGlyArgGlnArgAlaLeuSerAlaValAlaCysThrAlaSerGlyGlnArgIle 406
 QY 1198 CGCGCAGCGCGCTGCGCGCGCGCGCGCGCGCGCGGAGACCGCGCGCGTACGCG 1251
 DB 407 Arg-----ValProGlnAlaMetArgProGlnAlaGlyLeuArgGlnPro 421

RESULT 6
 AAB88340 standard; protein; 285 AA.
 XX
 AC AAB88340;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Human membrane or secretory protein clone P8C0053.
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes.
 XX
 OS Homo sapiens.
 XX
 PN EP1067182-A2.
 XX
 PD 10-JAN-2001.

XX 07-JUL-2000; 2000EP-00114090.
PF
XX
XX 08-JUL-1999; 99JP-00194179.
PR 11-JAN-2000; 2000JP-00118775.
PR 02-MAY-2000; 2000JP-00183766.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayaehi K;
XX WPI; 2001-09389/11.
DR N-PSDB; AAF93767.
XX
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development.
XX
PS Claim 1; SEQ ID NO 48; 609pp + Sequence Listing; English.

XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by AAB88317
CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
CC AAF92232 - AAF92235 which are used to isolate the cDNA sequences of the
CC invention. The invention also includes methods for the production of
CC antibodies directed against the proteins, and cDNA sequences, which can
CC be used in vaccines. The polynucleotide sequences can be used in gene
CC therapy. The polynucleotide sequences and the proteins they encode may be
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate secretory protein/membrane protein expression. The
CC nucleic acids and complementary sequences may also be used as DNA probes
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
CC and quantitate the presence of similar nucleic acid sequences in samples.
CC They may also be used to study the expression and function of secretory
CC proteins/membrane polypeptides and their role in metabolism. The
CC polypeptides may be used as antigens in the production of antibodies
CC against them and in assays to identify modulators (agonists and
CC antagonists) of expression and activity. The antibodies and antagonists
CC may also be used as therapeutic agents to down regulate expression and
CC activity. The antibodies may also be used as diagnostic agents for
CC detecting the presence of the polypeptides in samples (e.g. by enzyme
CC linked immunosorbent assay (ELISA)). Examples of diseases which may be
CC treated include rheumatoid arthritis and diabetes

XX
SQ Sequence 285 AA;

Alignment Scores:
Pred. No.: 1 94e-88 Length: 285
Score: 1409.50 Matches: 282
Percent Similarity: 98.9% Conservative: 0
Best Local Similarity: 98.9% Mismatches: 2
Query Match: 59.2% Indels: 2
DB: 4 Gaps: 1

US-10-699-035A-5 (1-1254) x AAB88340 (1-285)

QY 1 ATGCTCCCTGGAACGGCGCTGCGCTGAGCTTGCGGTGCGCGCGAGC 60
DB 1 MetLeuProTrrPThrIleuGlyLeuAlaLeuSerLeuArgLeuAlaArgSer 20
QY 61 GGGCGGAGACGGGCTCACCAGATAGCCCGGAGGAGCTGATGTTCTGCTGAGC 120
DB 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuAsp 40
QY 121 AGCTGACGACGGTCTCTCACTACGAGTCTCCCGGTTCCGAGATTGNGGGGACGTC 180
DB 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCACTGCGCTTGGGACACCGGGGCTGCTGCGCAGTCTGCTGACGCTGGCAGT 240
DB 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 GGGCCATACCGAGATTCCCTTGGCGCAGACACAGCTCGGGTGGAGCTGCCAGATGCG 300

DB 81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 100
QY 301 GTGGCTGCTTGGCCAGCGCATGGGTGACACCCACATGGCGCTGCGCTGATGCC 360
DB 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
QY 361 AAGAACAGCTGTTTGTCTGAAGCATCAGGTGCCCGGCGGAGGTCGCCAAAGTCTGGTG 420
DB 121 LysGluGlnLeuPheAlaGluAlaSerGlyAlaArgProGlyValProValValLeuVal 140
QY 421 TGGGTCAGATGGCGGCTCCAGGACCTGTGGGCGCCCGCATGACAGAGCTCAAGAC 480
DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGlnLeuAsp 160
QY 481 CTGGCGCTCACCGTTCATTGTGACGACCGCGCGAGGCACTTCTGGAAGCTGTGACGC 540
DB 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGlyLeuSerAla 180
QY 541 GTGGCTCAGCCCTGCGCGAGACCTGCACTTGTGACGTGAGATGACTGCAATC 600
DB 181 AlaAlaSerAlaProAlaGluAlaGlyHisLeuHisPheValAspAspLeuHisIle 200
QY 601 ATTGCTCAAGAGCTGAGGGGCTCCATTCTC--GGATGGCGCGCAGAGCTCCATGCC 657
DB 201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
QY 658 ACGGATCACGCTCCAGCGGCTTCCGCTGCGCTGCGCACCCCTGCTGACCGCAGACTCG 717
DB 221 ThrGluIleThrSerSerGlyPheArgLeuAlaTrpProProLeuLeuThrAlaAspSer 240
QY 718 GGCTNCTATGTGCTGAGACTGTGCGCCAGCGCCAGCGCGG-GGTGCAGAACGCGACCA 776
DB 241 GlyTyrTyrValLeuGlyLeuValProSerAlaGlnProGlyGlyValThrProAla 260
QY 777 GGTGCGAGGAAACGCGACGAGCTGATCTGGGCGCGGCTCGACCGGACAGACTAAGA 836
DB 261 AlaAlaArgGluArgGlyGlyLeuAspPheGlyArgProArgProGlyHisGlyLeuArg 280
QY 837 CGTGCGCTAGTGCC 851
DB 281 ArgGlyAlaSerAla 285
RESULT 7
ADY63045
ID ADY63045 standard; protein; 285 AA.
XX
XX
XX AC ADY63045;
XX
XX DT 02-JUN-2005 (first entry)
XX
XX DE Human clone PSEC0053 protein, SEQ ID 48.
XX
XX KW Gene therapy.
XX
XX OS Homo sapiens.
XX
XX PN EP1514933-A1.
XX
XX PD 16-MAR-2005.
XX
XX PF 07-JUL-2000; 2004EP-00027228.
XX
XX PR 08-JUL-1999; 99JP-00194179.
XX
XX PR 11-JAN-2000; 2000JP-00118775.
XX
XX PR 02-MAY-2000; 2000JP-00183766.
XX
XX PR 07-JUL-2000; 2000EP-00114090.
XX
XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayaehi K;
XX
XX DR WPI; 2005-203865/22.
XX
XX DR N-PSDB; ADY63044.

XX Novel isolated polynucleotide encoding human secretory proteins or
 PT membrane proteins, useful for examination and diagnosis of abnormality of
 PT human secretory proteins.
 XX
 XX Disclosure; SEQ ID NO 48; 1240bp; English.
 XX
 XX The present invention relates to novel human secretory proteins or
 CC membrane proteins, and their coding sequences. The present sequence is
 CC one such protein sequence. The coding sequences of the invention are
 CC useful for examination and diagnosis of abnormality of the human
 CC secretory proteins and in gene therapy methods. The coding sequences and
 CC proteins are useful as candidates for medicines or as target molecules
 CC for developing medicines. Antibodies against the proteins of the
 CC invention are useful for treating diseases that are associated with the
 CC proteins. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained from sequence information
 CC supplied by the European Patent Office.

XX
 XX Sequence 285 AA;

Alignment Scores:
 Pred. No.: 1,94e-88 Length: 285
 Score: 1409.50 Matches: 282
 Percent Similarity: 98.9% Conservative: 0
 Best Local Similarity: 98.9% Mismatches: 2
 Query Match: 59.2% Indels: 2
 DB: 9 Gaps: 1

US-10-699-035a-5 (1-1254) x ADY63045 (1-285)

QY 1 ATGCTCCCTGGAAGCGGCTGCGCTGAGCTTGGCGCTGCGCTGCGCGGAGC 60
 Db 1 MetLeuProThrPhrAlaLeuGlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
 QY 61 GGGCGGAGCGGCTGCGCTGAGCTTGGCGCTGAGCTTGGCGCTGAGCTTGGCGCTGAGC 120
 Db 21 GlyAlaGlyArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
 QY 121 AGCTGAGCGAGCTTCTCTCACTACGAGTTCTCCCGGCTTGGGAGTTTGGGGGAGCTG 180
 Db 41 SerSerAlaSerValSerHisSerIleThrGluPheSerArgValArgGluPheValGlyGlnLeu 60
 QY 181 GTGGCTCACTGCGCTTGGGAGCAGCGGAGCGCTGCGCTGAGCTTGGCGCTGAGCTGAGC 240
 Db 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuAlaGlyAlaSerLeuValHisValGlySer 80
 QY 241 CGGCTTACACCGAGTTCCTCCCTTGGCGGAGCAGCTGCGCTGAGCTTGGCGCTGAGCTGAGC 300
 Db 81 ArgProIleThrGluPheProPheGlyGlnHisSerSerGlyGlyAlaAlaGlnAspAla 100
 QY 301 GTGGCGCTTCTGCGGAGCGAGTGGGAGCAGCTGAGCTTGGCGCTGAGCTTGGCGCTGAGC 360
 Db 101 ValAlaGlySerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValIleVal 120
 QY 361 AAGGAACAGCTGTTGGCTGAGCATGAGTGGCGGAGCGGAGCTGCGGAGCTGAGCTGAGC 420
 Db 121 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProIleValLeuVal 140
 QY 421 TGGGTCACAGATGGCGGCTTGGCGGAGCAGCTTGGCGGAGCTTGGCGGAGCTGAGC 480
 Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProProMetGlnGluLeuLysAsp 160
 QY 481 CTGGGCGTACAGCTGTTGCTTGGCGGAGCAGCTTGGCGGAGCTTGGCGGAGCTGAGC 540
 Db 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 180
 QY 541 GGTGCTCAGCGCTTGGCGGAGCAGCTTGGCGGAGCTTGGCGGAGCTGAGC 600
 Db 181 AlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspValAspLeuHisIle 200
 QY 601 ATTGTTCAGAGCTGAGGCTTCCATTCTC--GCATGCGCGGAGCAGCTTCCATTGCC 657

Db 201 IleValGlnLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
 QY 658 ACGAGATCAGCTTCCAGCGGCTTCCGCTGAGCTGAGCTTCCGCTGAGCTTCCGCTGAGCTGAGC 717
 Db 221 ThrGluIleThrSerSerGlyPheArgLeuAlaTrpProProLeuLeuThrAlaAspSer 240
 QY 718 GGTCTACTATGCTGAGCTGAGTGGCTGCGGAGCTTGGCGGAGCTTGGCGGAGCTGAGC 776
 Db 241 GlyTyrTyrValLeuGlyLeuValProSerAlaGlnProGlyGlyCysIleThrProAla 260
 QY 777 GTGGCGAGGAGCAGCGGAGCTGAGCTTGGCGGAGCTTGGCGGAGCTTGGCGGAGCTGAGC 836
 Db 261 AlaAlaArgGlyArgHisIleIleuAspLeuGlyArgProArgProGlyHisGlyLeuArg 280
 QY 837 CGTGGCGCTAGTGCC 851
 Db 281 ArgGlyAlaSerAla 285

RESULT 8
 AAB42581
 ID AAB42581 standard; protein; 299 AA.
 XX
 AC AAB42581;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2345 polypeptide sequence SEQ ID NO:4690.
 XX
 KW Human; open reading frame; ORFX; detection; cytoskeletal; hepatotropic;
 KW vulnary; antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antipneumatic; antihydroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; inflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 KW
 XX
 OS Homo sapiens.
 OS
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US008621.
 XX
 PF 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.
 PA
 XX
 PI Shimkete RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR
 DR N-PSDB; AAC76790.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 PT
 XX
 PS Claim 11; Page 3880-3881; 5507bp; English.
 PS
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytoskeletal; hepatotropic; vulnary;
 CC antiparkinsonian; neurotropic; neuroprotective; osteopathic;

CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antifungal; antineumatic; antithyroid; and antianaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORX-associated disorder. The nucleic acids can be used to express ORX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 299 AA;

Alignment Scores:

Pred. No.:	6.32e-86	Length:	299
Score:	1373.00	Matches:	282
Percent Similarity:	94.6%	Conservative:	0
Best Local Similarity:	94.6%	Mismatches:	8
Query Match:	57.7%	Indels:	8
DB:	3	Gaps:	3

US-10-699-035a-5 (1-1254) x AAB42581 (1-299)

QY 3 GCTCCCGGAGAGGGGCTGCGCTGCGCTTGGCGGCTTGGCGGCGGAGCGG 62
 Db 9 AAlaProLeuAspGlyAlaArg-LeuAlaLeuSerLeuAlaG-TyrArgTyrPArg-----T 26
 QY 63 CGCGAGGCGGCTCCAGCATCAGATCAGCCCGGAGGAGCATGATGTTCTCTGAGCAG 122
 Db 26 hrProAspCys--ProProAlaSerAlaProGlu---AspLeuMetPheLeuLeuAspSe 44
 QY 123 CTGACCGCAGCGTCTCTACTACGAGTTCTCCGGGTTGGGAGTTTGTGGGCGACGTG 182
 Db 44 rSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGluGluLeuVal 64
 QY 183 GCGTCACATGCGCCCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
 Db 64 lAlaProLeuProLeuAlaPro**AlaLeuAlaArgAlaSerLeuValHisValGlySerAr 84
 QY 243 GGCATACACGAGTTCCCTTGGCGGCGGAGCATGCGGAGTGGAGGCGGAGTGGCGGT 302
 Db 84 gProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAlaVa 104
 QY 303 GCGTCCTTGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 362
 Db 104 lArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAlaLy 124
 QY 363 GGAAGACGCTGTTGCTGGAAGCATGAGTCCCGGCGGAGGCGGCGGCGGCGGCGG 422
 Db 124 sGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProGlyValLeuValTr 144
 QY 423 GGTGACAGATGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 482
 Db 144 pValThrAspGlyGlySerSerAspProValGlyProPrometGlnGluLeuLysAspLe 164
 QY 483 GGGCGCTACCGGTGTCATTGTGACGACCGGCGGAGCGGAGCGGAGCGGAGCGGAG 542
 Db 164 uGlyAlaThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAlaAl 184
 QY 543 TGGCTCAGCGGCGGCGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 602
 Db 184 aAlaSerAlaProAlaGlnLysHisLeuHisPheValaLysAspAspLeuHisIleLe 204
 QY 603 TGTCCAAAGAGTGAAGGAGCTCATTTCTC--GCGATGCGGCGGAGCGGAGCGGAG 659
 Db 204 eValGlnGluLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisIleAla 224

QY 660 GGAATCATCCTCCAGCGGCTTCCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGG 719
 Db 224 rGluLeuThrSerSerGlyPheArgLeuAlaTyrProProLeuLeuThrAlaAspSerG 244
 QY 720 CTACTATGCTGAGAGCTGAGTGGCCAGCGGCGGAGGCTGCAAGACGCGACGACT 779
 Db 244 YTrpTyrValLeuGluLeuValProSerAlaGlnProGlyAlaAlaArgGlnGlnLe 264
 QY 780 GCCAGGAGACCGCACGAGCTGATCTGGGCGGCGGCTCGACCGGAGCACGAGTACGAG 839
 Db 284 uProGlyAsnAlaThrAspTyrPleTyrAlaGlyLeuAspProAspThrAspTyrAspVa 284
 QY 840 GCGGTATGCTGAGTCAACGTCGCGCTCTGAGCGCCGACATC 885
 Db 284 lAlaLeuValProGlnSerAsnValArgLeuLeuArgProGlnLe 299

RESULT 9
 AAB87344
 ID AAB87344 standard; protein, 215 AA.
 XX
 AC AAB87344;
 XX
 DT 22-MAY-2001 (first entry)
 XX
 DE Human gene 3 encoded secreted protein HMTB078, SEQ ID NO:85.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes;therosclerosis;
 KW cardiovascular disorder; angioygenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnery; cell culture;
 KW chemotaxis; food additive; binding partner identification.
 XX
 OS Homo sapiens.
 XX
 PN WO200118022-A1.
 XX
 PD 15-MAR-2001.
 XX
 PF 31-AUG-2000; 2000WO-US024008.
 XX
 PR 03-SEP-1999; 99US-0152315P.
 PR 03-SEP-1999; 99US-0152317P.
 XX
 PA (HUMA-) HUMAN GENOME SCT INC.
 XX
 PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
 PI Sopet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
 PI Moore PA, Shi Y, Wei Y, Florence KA;
 XX
 DR WPT, 2001-203081/20.
 DR N-PSDB; AAF91860.
 XX
 PT Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers.
 XX
 PS Claim 11; Page 532-533; 607P; English.
 XX
 AAFA1858-AAF91929 represent cDNAs corresponding to 52 human secreted
 CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
 CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 52 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of

CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
 CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiotensin disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the proteins mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention
 CC
 XX
 SQ Sequence 215 AA;

Alignment Scores:

Pred. No.:	5, 95e-65	Length:	215
Score:	1067.50	Matches:	213
Percent Similarity:	99.1%	Conservative:	0
Best Local Similarity:	99.1%	Mismatches:	1
Query Match:	44.9%	Indels:	1
DB:	4	Gaps:	1

US-10-699-035A-5 (1-1254) x AAB87344 (1-215)

QY 1 ATGCTCCCTGAGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCGAGC 60
 1 MetLeuProTrpPhrAla***GlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaGser 20
 QY 61 GCGCGCGAGCGCGCTGCGCGAGCATGAGCCCGCGAGGAGCCTGATGTTCTGCTGAGC 120
 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
 QY 121 AGCTGAGCGAGCTGCTGCTGAGTCTCCGCGGTTGCGAGTTGCGGCGAGCTG 180
 41 SerSerAlaSerValSerHisTrpGluPheSerAlaArgGluPheValGlyGlnLeu 60
 QY 181 GTGCTCCAGTGGCGCGAGCGCGAGCGCGCTGCTGCGAGTGTGTCAGTGGCGAGT 240
 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
 QY 241 CGGCGATACACGAGTTCCTGCGCGAGCGAGCATGCTGCGTGGAGCTGCCAGATGCG 300
 81 ArgProTrpThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 100
 QY 301 GTGCGGTCTTTCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC 360
 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValIYrAla 120
 QY 361 AAGGAACAGCTGTTTCTGAGCATGAGGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCTG 420
 121 LySGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyAlaProLySerValLeuVal 140
 QY 421 TGGGTGACAGATGGCGCTGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC 480
 141 TrpValThrAspIYlGlySerSerAspProValGlyProPheMetGlnGluLeuLysAsp 160
 QY 481 CTGGGCGTACCGGTTCATTTGTCAGACCGCGCGAGGAGAACTTCTGAGCGTGTGAGC 540
 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 180
 QY 541 GCTGCTCAGCGCGCTGCGAGGAGCATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGC 600
 181 AlaAlaSerAlaProAlaGluLysHisLeuHisPheValAspValAspAspLeuHisIle 200
 QY 601 ATTGTCAAGAGCTGAGGCGCTGCAATTC--GCGATGCGGCGG 642

DB 201 ILeValGlnGluLeuArgGlySerIleLeuAspAlaMetArgPro 215
 RESULT 10:
 ABG65347
 ID ABG65347 standard; protein; 215 AA.
 XX
 AC ABG65347;
 XX
 DT 27-AUG-2002 (first entry)
 DE
 XX
 XX Human albumin fusion protein #2022.

XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antifertility; antiinflammatory; anticancer;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.
 XX
 OS Homo sapiens.
 OS Synthetic.
 PN WO200177137-A1.
 PD 18-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-US011988.
 PR 12-APR-2000; 2000US-0229358P.
 PR 25-APR-2000; 2000US-0199384P.
 PR 21-DEC-2000; 2000US-0256931P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Haseltine WA;
 XX
 DR WPI; 2002-010886/01.

PT New fusion protein for treating disease e.g. diabetes comprises an
 PT albumin fused to a therapeutic protein.
 XX
 XX
 PS Claim 1; Page 1935; 2102pp; English.

CC The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA, also known as human serum
 CC albumin, HSA). The proteins are useful for treating a disease or disorder
 CC that may be modulated by therapeutic protein X. The albumin extends the
 CC shelf-life of protein X, and may increase its biological in vitro/in vivo
 CC activity. The protein is useful for treating and diagnosing disorders
 CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
 CC disease, ulcerative colitis), immune disorders (e.g. acquired
 CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
 CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,
 CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
 CC schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).
 CC ABG65326-ABG65518 represent albumin fusion proteins of the invention
 XX
 SQ Sequence 215 AA;

Alignment Scores:

Pred. No.:	5, 95e-65	Length:	215
Score:	1067.50	Matches:	213
Percent Similarity:	99.1%	Conservative:	0
Best Local Similarity:	99.1%	Mismatches:	1
Query Match:	44.9%	Indels:	1
DB:	5	Gaps:	1

US-10-699-035A-5 (1-1254) x ABG65347 (1-215)

QY 1 ATGCTCCCTGAGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCGAGC 60

Db 1 MetLeuProTrpThrAla***GlyLeuAlaLeuSerLeuAlaGlyLeuAlaLeuAlaArgSer 20
 QY 61 GCGCGGAGCGCGGCTCCACAGCATCAGCCCGGAGGGAGCTGATGTTCTGCTGAGAC 120
 Db 21 GlyAlaGluValGlyProProAlaSerAlaProAlaGlyAspLeuMetPheLeuLeuAsp 40
 QY 121 AGCTCAGCCAGCGTCTCTCAGTACGAGTTCCTCCGAGTTCGGAGTTCGGGAGCTG 180
 Db 41 SerSerAlaSerValSerHisIerYrGluPheSerArgValArgGluPheValGlyGlnLeu 60
 QY 181 GTGGCTCAGTCCCTGGGAGCGGAGGCGCTGGCGGAGCTGCTGAGCTGGGAGCT 240
 Db 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuAlaGlyAlaSerLeuValHisValGlySer 80
 QY 241 CGGCGATACACGAGTTCCTCCGAGCGAGCATCAGCTCGGAGTTCGGAGTTCGGAGTGC 300
 Db 81 ArgProYrTrpThrGluPheProPheGlyGlnHisSerSerGlyGluAlaGlnAspAla 100
 QY 301 GTGGCTGCTTCTGCGGAGCGGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 360
 Db 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisSThrGlyLeuAlaLeuValTyrAla 120
 QY 361 AAGGACAGCTGTTTGTGAGATCAGTTCGCGGAGGAGGAGTTCGCGGAGTTCGAGTTC 420
 Db 121 LysGluGlnLeuPheAlaGluAlaSerGlyAlaArgProGlyValProLysValLeuVal 140
 QY 421 TGGGAGACAGATGAGGAGGCTCCAGCGAGCTGGGAGGCGCGGAGTTCGAGTTCGAGTTC 480
 Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGlnLeuLysAsp 160
 QY 481 CTGGGCGTACCGGCTTCATTGTGACAGCGGCGGAGGAGTTCGAGTTCGAGTTCGAGTTC 540
 Db 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGlnLeuSerAla 180
 QY 541 GCTGCTCAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
 Db 181 AlaAlaSerAlaProAlaGlnLysHisIleValHisPheValAspValAspAspLeuHisIle 200
 QY 601 ATTGTCAGAGAGCTGAGGAGGCTCCATTCTC--GGCATCGGCGCG 642
 Db 201 IleValGlnGlnLeuArgGlySerIleLeuAspAlaMetArgPro 215
 RESULT 11
 ADL78614
 ID ADL78614 standard; protein; 215 AA.
 AC ADL78614;
 XX
 XX 20-MAY-2004 (first entry)
 DT
 XX
 XX Albumin fusion protein related therapeutic protein X, SEQ ID No 2096.
 DE
 XX albumin fusion protein; cytostatic; antinaemic; antiarthritic;
 XX antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;
 XX antiparasitic; antibacterial; osteoplastic; dermatological; antiout;
 XX immunomodulator; antiarrhythmic; cardiac; neotropic; antidiabetic;
 XX nephrotoxic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
 XX antidiabetic; anabolic; hypertensive; vulnary; gene therapy; cancer;
 XX reproductive system disorder; therapeutic protein.
 OS
 XX Unidentified.
 XX
 XX US2004010134-A1.
 XX
 XX 15-JAN-2004.
 XX
 XX 12-APR-2001; 2001US-00833245.
 XX
 XX 12-APR-2000; 2000US-0229358P.
 XX 25-APR-2000; 2000US-0199384P.
 XX 21-DEC-2000; 2000US-0256931P.
 XX

PA (ROSE/) ROSEN C A.
 PA (HASE/) HASELTINE W A.
 XX
 PI Rosen CA, Haseeltine WA;
 XX
 DR WPI: 2004-090519/09.
 XX
 XX
 PS Disclosure: SEQ ID NO 2096; 279pp: English.

The invention relates to a novel albumin fusion protein. The invention further relates to a composition comprising the albumin fusion protein and a pharmaceutical carrier; a kit comprising the composition of the albumin fusion protein formula; a method of treating a disease or disorder in a patient comprising the step of administering the albumin fusion protein; a method of treating a patient with a disease or disorder that is modulated by Therapeutic protein: X, or its fragment or variant; a method of extending the shelf life of Therapeutic protein: X, or its fragment or variant; a nucleic acid molecule comprising a polynucleotide sequence encoding the albumin fusion protein; a vector comprising the nucleic acid molecule of the albumin fusion protein; and a host cell comprising the nucleic acid molecule of the albumin fusion protein. The albumin fusion protein and its compositions have the following activities: cytostatic, antinaemic, antiarthritic, antiasthmatic, anti-HIV, immunosuppressive, antiinflammatory, antiparasitic, antibacterial, osteoplastic, dermatological, antitumor, immunomodulator, antiarrhythmic, cardiac, neotropic, antidiabetic, nephrotoxic, uropathic, neuroprotective, antiParkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive, and vulnary. The albumin fusion protein nucleic acid may be used in gene therapy to treat disorders. The albumin fusion protein is useful for diagnosing, treating, preventing or ameliorating diseases or disorders comprising indication: Y. The diseases or disorders include: cancer (e.g. leukemia, colon, bone, breast, liver or lung cancer), immune or hematopoietic diseases (e.g. anemia, Hodgkin's disease, acute lymphocytic leukaemia, multiple myeloma, arthritis, asthma, AIDS, autolymphoma disease, inflammatory bowel disease, psoriasis or Lyme disease), reproductive system disorders (e.g. prostatic, inguinal hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-leydig tumours), musculoskeletal diseases (e.g. giant cell tumours, Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy or cachexia), cardiovascular disease (e.g. Thrombomas, heart disease, arrhythmia, cardiac arrest, heat valve disease, hypernatraemia or hyponatraemia), mixed foetal diseases (e.g. foetal alcohol syndrome, Down's syndrome, Patou syndrome, Turner's syndrome, Apert syndrome or Tay-Sachs disease), excretory diseases (e.g. urinary incontinence, urinary tract infections or renal disorders), neural or sensory disease (e.g. Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis, cerebellar ataxia, attention deficit disorder, autism or obsessive compulsive disorder), respiratory diseases (e.g. emphysema, lung cancer or occupational lung disease), endocrine diseases (e.g. diabetes, Addison's disease or glomerulonephritis), digestive diseases (e.g. portal hypertension, irritable bowel disease, gastric atrophy or pancreatitis) or connective tissue or epithelial diseases (e.g. Crohn's disease, scleroderma, wound healing or epidermolysis bullosa). This sequence CC represents a therapeutic protein X relating to the albumin fusion protein of the invention. The sequence listing data for this specification was downloaded from the USPTO website.

XX
 XX SQ Sequence 215 AA;
 XX
 XX Alignment Scores:
 Pred. No.: 5,956-65
 Score: 1067.50
 Percent Similarity: 99.1%
 Best Local Similarity: 99.1%
 Query Match: 44.9%
 DB: 8
 Gaps: 1
 US-10-699-035A-5 (1-1254) x ADL78614 (1-215)

QY 1 ATGCTCCCTGAGACGGCGCTCGGCTGAGCTTGCGGCTGGCGCTGGCGGAGC 60
DB 1 MetleuProtrpThrAla**GlyleuAlaIleuSerIleuArgIleuAlaIargSer 20
QY 61 GGGCGGAGCGGGTCCACAGATACAGCCCGGAGGGGAGCTGATGTTCCGCTGAGC 120
DB 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspIleuIleuAsp 40
QY 121 AGCTGAGCAGCGCTCTCTCACTACGAGTTCTCCGGGCTCGGAGTTTGCGGAGCTG 180
DB 41 SerSerAlaSerValSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 60
QY 181 GTGGCTCCACTGCCCCCTGGGCAACCGGGGCGCTGCTGCTGCTGCTGCTGCTGCT 240
DB 61 ValAlaProIleuProIleuGlyThrGlyAlaIleuArgAlaSerIleuValHisValGlySer 80
QY 241 CGGCGATACACGAGTTCCCTTGGCGGAGCTGCGGCTGAGGCTGCGGAGCTGCGGAG 300
DB 81 ArgProIleuThrGluPheProIleuGlyGlnHisSerSerGlyGlnAlaIleuAla 100
QY 301 GTGGCTGCTTGGCGGAGCGATGAGTGAACACCACTGCTGCTGCTGCTGCTGCTGCT 360
DB 101 ValAlaGluAlaSerAlaGlnArgMetGlyAspThrHisThrGlyIleuAlaIleuValIleuAla 120
QY 361 AAGGAACAGCTGTTGCTGAGCATGAGTGCCTGCGGCGGAGGCTGCTGCTGCTGCTG 420
DB 121 LysGluGlnIleuPheAlaGlnIleuAlaSerGlyAlaArgProGlyValProIleuValIleuVal 140
QY 421 TGGGTGACGATGAGCGGCTGCGGAGCGCTGCTGCGGCGGAGGCTGCTGCTGCTGCTG 480
DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProIleuIleuIleuValIleuVal 160
QY 481 CTGGGCGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 161 LeuGlyValThrValPheIleuAlaSerIleuThrGlyArgGlyAsnIleuIleuSerAla 180
QY 541 GGTGCTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 181 AlaAlaSerAlaProAlaGlnIleuSerIleuAlaSerValAspValAspIleuHisIleu 200
QY 601 ATTTGTCAGAGCTGAGGCGCTTCATTC--GCATGCGGCGG 642
DB 201 IleValGlnIleuArgGlySerIleuAspAlaMetArgPro 215
RESULT 12
AAB87418
ID AAB87418 standard; protein; 242 AA.
AC AAB87418;
XX
XX
DT 22-MAY-2001 (first entry)
XX
XX
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
XX
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiodysplasia; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; cell culture;
KW chemotaxis; food additive; binding partner identification.
OS Homo sapiens.
PN WO200118022-A1.
XX
XX
PD 15-MAR-2001.
XX
XX
PF 31-AUG-2000; 2000MO-US024008.
XX

PR 03-SEP-1999; 99US-0152315P.
PR 03-SEP-1999; 99US-0152317P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ni J, Baker KP, Birse CE, Piscella M, Komatsuji GA, Rosen CA;
PI Soppet DR, Young PE, Edner R, Duan DR, Olsen HS, Latleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
XX
DR WPI; 2001-203081/20.
XX
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
PS Disclosure; Page 18; 607pp; English.
XX
XX
CC AAF9185E-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiodysplasia, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX
SQ Sequence 242 AA;
XX
XX
Alignment Scores:
Pred. No.: 6,09e-65 Length: 242
Score: 1067.50 Matches: 213
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 44.9% Indels: 1
DB: 4 Gaps: 1
US-10-699-035a-5 (1-1254) x AAB87418 (1-242)
QY 1 ATGCTCCCTGAGACGGCGCTCGGCTGAGCTTGCGGCTGGCGCTGGCGGAGC 60
DB 1 MetleuProtrpThrAla**GlyleuAlaIleuSerIleuArgIleuAlaIargSer 47
QY 28 MetleuProtrpThrAla**GlyleuAlaIleuSerIleuArgIleuAlaIargSer 47
DB 28 MetleuProtrpThrAla**GlyleuAlaIleuSerIleuArgIleuAlaIargSer 47
QY 61 GGGCGGAGCGGGTCCACAGATACAGCCCGGAGGGGAGCTGATGTTCCGCTGAGC 120
DB 48 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspIleuIleuAsp 67
QY 121 AGCTGAGCAGCGCTCTCTCACTACGAGTTCTCCGGGCTCGGAGTTTGCGGAGCTG 180
DB 68 SerSerAlaSerValSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 87
QY 181 GTGGCTCCACTGCCCCCTGGGCAACCGGGGCGCTGCTGCTGCTGCTGCTGCTGCT 240

Db 88 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 107
Qy 241 CGGCCATACACCGAGTTCCCTTGGCCAGCAGCTCGGGTGAAGCTGCCAGGATGCG 300
Db 108 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 127
Qy 301 GTGGGTGCTTTGGCCAGCAGATGGGTGAACCCACACTGGCTGGCGCTGATGCC 360
Db 128 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 147
Qy 361 AAGGAACAGCTGTTTGTGAAGCATCAGTGGCCGGCCAGGGGTGCCAAAGTGTGGTG 420
Db 148 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValAlaProValLeuVal 167
Qy 421 TGGGTGACAGATGGCGGCTCCAGCGACCTTGGGCCCCCAGATGACAGAGTCAAGAC 480
Db 168 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGluLeuLysAsp 187
Qy 481 CTGGGCGCTCACCGTGTATTGTTCAGCACCGGCCAGGCACTTCTGGAGCTGTACACC 540
Db 188 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 207
Qy 541 GCTGCTCAGGCCCCCGGAGAGACCTGACCTTGGAGCGTGGATGACCTGCACATC 600
Db 208 AlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspValAspAspLeuHisIle 227
Qy 601 ATGTGTCAGAGCTTGAGGGGCTCATTTCTC--GGATGCGGCGC 642
Db 228 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgPro 242
RESULT 13
AAB87424
ID AAB87424 standard; protein; 226 AA.
XX
AC AAB87424;
XX
DT 22-MAY-2001 (first entry)
XX
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnerrary; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
OS Homo sapiens.
XX
PN WO200118022-A1.
XX
PD 15-MAR-2001.
XX
PF 31-AUG-2000; 2000WO-US024008.
XX
PR 03-SEP-1999; 99US-015231SP.
XX
PR 03-SEP-1999; 99US-0152317P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsuolis GA, Rosen CA,
PI Soper DR, Young PE, Edner R, Duan DR, Olsen HS, Lafleur DW,
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
DR WPI, 2001-203081/20.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.

XX
PS Disclosure; Page 18; 607pp; English.
XX
AA AAF9158-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87413-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
CC
XX
SQ Sequence 226 AA;
XX
Alignment Scores:
Pred. No.: 1.44e-64 Length: 226
Score: 1062.00 Matches: 210
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 44.6% Indels: 0
DB: 4 Gaps: 0
US-10-699-035A-5 (1-1254) x AAB87424 (1-226)
Qy 1 ATGTCCCTCCGAGCGGCTGAGCTGAGCTTGGAGCTTGGCGGAGC 60
Db 14 MetLeuProTyrThrAlaLeuGlyLeuAlaLeuSerLeuAlaLeuAlaArgSer 33
Qy 61 GAGCGGAGCGCGGTCCACACGATCAGCCCCCGAGGGACCTGATTTCTGTGAC 120
Db 34 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 53
Qy 121 AGCTCAGCAGCGTCTCCATGAGTTCCTCCGGGTTCGGGAGTTTGGGGCAGCG 180
Db 54 SerSerAlaSerValSerHisIleTyrGluPheSerArgValArgLysPheValGlyGlnLeu 73
Qy 181 GTGGCTCAGTCCCTTGGGACCGGAGCGGCTGCTGACGTTGGTGCAGTGGGACGT 240
Db 74 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 93
Qy 241 CGGCCATACACCGAGTTCCCTTGGCCAGCAGCTCGGGTGAAGCTGCCAGGATGCG 300
Db 94 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 113
Qy 301 GTGGGTGCTTTGGCCAGCAGATGGGTGAACCCACACTGGCTGGCGCTGATGCC 360
Db 114 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 133
Qy 361 AAGGAACAGCTGTTTGTGAAGCATCAGTGGCCGGCCAGGGGTGCCAAAGTGTGGTG 420
Db 134 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValAlaProValLeuVal 153

QY 421 TGGGTGACAGATGGCGGCTTCACGCCACCTGTGGGCCCCCATGACGAGACTCAAGAGAC 480
DB 154 TTPVATHTHTAspGlyGlySerSerAspProValGlyProProwMetGlnGluLeuLysAsp 173
QY 481 CTGGGGCTACCGGTGTTCAATTGTCAGCACCGCGGAGCACTCTCTGAGCTGTACGCC 540
DB 174 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 193
QY 541 GGTGCTTCAAGCCCTGCGGAGCAAGCACTGACCTTTGTGAGCGTGATGACCTGCACATC 600
DB 194 AlaAlaSerAlaProAlaGlnGlySerIleLeuHisPheValAspValAspAspLeuHisIle 213
QY 601 ATTGTTCAGAGCTGAGGGGCTTCATTCTC 630
DB 214 ILeValGlnGluLeuArgGlySerIleLeu 223
RESULT 14
AAB87419
ID AAB87419 standard; protein, 186 AA.
AC AAB87419;
XX
XX
DT 22-MAY-2001 (first entry)
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:160.
XX
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiotensin disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery; cell culture;
KW chemocaxis; food additive; binding partner identification.
XX
XX
OS Homo sapiens.
XX
XX
PN MO200118022-A1.
XX
XX
PD 15-MAR-2001.
XX
XX
PF 31-AUG-2000; 2000MO-US024008.
XX
XX
PR 03-SEP-1999; 99US-015231SP.
PR 03-SEP-1999; 99US-0152317P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
PI Soppet DR, Young PE, Ebnier R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA.
XX
XX
DR WPI; 2001-203081/20.
XX
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
XX
PS Disclosure; Page 18; 607p; English.
XX
XX
CC AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87419-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis and treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental

CC abnormalities, hematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiotensin disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX
SQ Sequence 186 AA;
XX
XX
Alignment Scores:
Pred. No.: 5 33e-54 Length: 186
Score: 908.00 Matches: 156
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 38.2% Indels: 0
DB: 4 Gaps: 0
US-10-699-035a-5 (1-1254) x AAB87419 (1-186)
QY 472 GCTCTGACAGTGGGGGGGCGCCACAGGGGTGCTGAGCCCGCATCTGTACACACAGCA 413
DB 1 AlaProAlaTrpGlyGlyProGlnGlyArgTrpSerArgHisLeuSerProThrProAla 20
QY 412 CTTTGGGACCCCTGGCGGGGACCTGATGCTTCAGCAACAGCTTCTTGGCATTA 353
DB 21 LeuTrpAlaProLeuAlaGlyHisLeuMetLeuGlnGlnThrAlaValProTrpHisArg 40
QY 352 CGAGGCGCAGCGCAGTGGGTGGGTGACCCATCGGTGGGCGAGAGACGACCGACCTCT 293
DB 41 ProAlaProGlyGlnCysGlyHisProCysAlaGlyGlnLysHisAlaProHisPro 60
QY 292 GGGCAGCCTCACCCAGCTGTGCTGCGGAGAGGGGAACCTCGGTATGACCGACCTGCC 233
DB 61 GlyGlnProHisProSerCysHisAlaGlyArgGlyThrArgCysMetAlaAspCysPro 80
QY 232 CGTGACACAGACTGACGACGACGCGGCTCCGAGGAGACGTGTATGACCGACCTGCC 173
DB 81 ArgAlaProAspTrpHisAlaGlyProArgCysProGlyAlaValGluProProAlaAla 100
QY 172 CCACAAACCTCCGGAACCGGAGAACTGTGTAGAGAGCGGTGCTGAGCTGTCCAGCA 113
DB 101 ProGlnThrProGlnProGlyArgThrArgSerGluArgArgTrpLeuSerCysProAla 120
QY 112 GGAACATCAGTCCCTCGGGGGCTGATGTCTGTGACCGGCGCTCCGCGCTCCGG 53
DB 121 GlyThrSerGlyProLeuGlyGlyLeuMetLeuValAspArgAlaProArgArgSerAla 140
QY 52 CCAGGCGCAGCGCAGACTCAGGGCGGCGGAGCGCGCTCCAGGGAGCA 2
DB 141 ProAlaProAlaAlaSerSerGlyProGlyArg***ProSerArgGlyAla 157
RESULT 15
AAB32500
ID AAB32500 standard; protein, 180 AA.
AC AAB32500;
XX
XX
DT 24-MAR-2003 (first entry)
XX
XX
DE Human von Willebrand Factor A (VA) domain.
XX
XX

KW Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
 KW extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
 KW gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 1..2
 FT /note= "Encoded by GGG"
 FT Misc-difference 179..180
 FT /note= "Encoded by CTC"
 XX
 XX WO20028184-A1.
 XX
 XX PD 07-NOV-2002.
 XX
 XX PF 02-MAY-2002; 2002WO-AU000542.
 XX
 XX PR 02-MAY-2001; 2001AU-00004701.
 XX
 XX PA (MURD-) MURDOCH CHILDRENS RES INST.
 XX
 XX PI Bateman JF, Fitzgerald DJ;
 XX
 XX DR WPI; 2003-111873/10.
 XX
 XX DR N-PSDB; AAD50397.
 XX
 XX PT New isolated Willebrand Factor A-Related Protein polypeptide useful for
 PT the manufacture of a medicament in the treatment of a disease condition
 PT of the extracellular matrix, in particular arthritis.
 XX
 XX PS Claim 7, Page 72-73; 103pp; English.
 XX
 XX CC The invention relates to Willebrand Factor A domain related-protein
 CC (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
 CC superfamily of extracellular matrix (ECM) proteins. WARP is used as an
 CC molecular marker, used for detecting a loss of ECM integrity in an animal
 CC subject, monitoring repair, regeneration or other disease processes in an
 CC animal subject and detecting a disease condition or a propensity for the
 CC development of a disease condition in an animal subject. The invention is
 CC useful for the manufacture of a medicament in the treatment of a disease
 CC condition of the ECM. The disease condition involves the cartilage, and
 CC is preferably arthritis. The invention is also used in gene therapy. The
 CC present sequence is human VA domain
 XX
 XX SQ Sequence 180 AA;
 Alignment Scores:
 Pred. No.: 6,21e-54 Length: 180
 Score: 907.00 Matches: 179
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 38.1% Indels: 0
 DB: 6 Gaps: 0
 US-10-699-035A-5 (1-1254) x AAE32500 (1-180)
 QY 94 CGAGGGGACCTGATGTTCTGCTGACAGCTCAGCCAGCGTCTCTCACTACGAGTTCTCC 153
 Db 1 ArgGlyAspLeuMetLeuAspSerSerAlaSerValSerHisTyrGluPheSer 20
 QY 154 CGGGTTCCGGAGATTGTGGGGCAGCTGTGCTCCACTGCCCTGGGACCGGGGCCCTG 213
 Db 21 ArgValArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeu 40
 QY 214 CGTCCAGTCTGGTGCAGCTGGGCACTGCGCCATACCGAGTTCCTCCCTTCGGCCAGCAC 273
 Db 41 ArgAlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHis 60
 QY 274 AGCTCGGAGGAGCTGCCAGAGTGGCGGTGCGGTCTTCCAGCGAGCGAGGTGACACC 333
 Db 61 SerSerGlyGluAlaIleGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThr 80

QY 334 CAACTGCGCTGGCGGCTGTGCTATAGCCAGAGAACGCTGTTGCTGAAGCATCAGGTGCC 393
 Db 81 HisThrGlyLeuAlaLeuValTyrAlaGlyGlnLeuPheAlaGluAlaSerGlyAla 100
 QY 394 CGGCCAGGGGTGCCAAAGTGTGTTGGGTGACAGATGAGCGGCTCCAGGACCCCTGTG 453
 Db 101 ArgProGlyValProLysValLeuValIrrpValThrAspGlyGlySerSerAspProVal 120
 QY 454 GGGCCCCCATGACAGAGCTCAAGGACCTGGGCGTCAACCGTTCATTGTGACGACCGGC 513
 Db 121 GlyProMetGlnGluLeuLysAspLeuGlyValThrValPheIleValSerThrGly 140
 QY 514 CGAGCAACTTCTCGGAGCTGTCAAGCGCTGCTGACCGCTCCAGCCCTCCGAGAAAGCACTGAC 573
 Db 141 ArgGlyAsnPheLeuGluLeuSerAlaAlaAlaSerAlaProAlaGluLysHisLeuHis 160
 QY 574 TTTGTGACGTGGATGACTGACATCATTTGTCCAGAGCTGAGGGGCTCCATTCTC 630
 Db 161 PheValAspValAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleLeu 179

Search completed: February 13, 2006, 13:23:41
 Job time : 228.052 secs

Db 1 MetLeuProTPrThAlaLeuGlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
QY 61 GCGCGGAGCGCGCTCCACCAAGCATAGCCCCCGAGGGAGCTGATGTTCTCGTGAGAC 120
Db 21 GlyAlaGlnArgGlyProProAlaSerAlaProAlaGlyAspLeuMetPheLeuLeuAsp 40
QY 121 AGCTCAGGCGAGGCTCTCTACACTACAGATTCTCCGGGATTCCGGAGTTTGGGGGACGCTG 180
Db 41 SerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlnPheValGlyGlnLeu 60
QY 181 GTGGCTCACTGCCCCCTGGGGACCGGGGGCCCTGGCGCCAGCTGAGTGCATGCGGACGT 240
Db 61 ValAlaAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGCCATACACCGAGTTCCCTTCGCGCAGCAGCTCGGGTGAAGAGCTGCGCCAGGATGCG 300
Db 81 ArgProTyrThrGlnPheProPheGlyGlnHisSerSerGlyGlnAlaAlaGlnAspAla 100
QY 301 GTGGCTGCTTTCGCCAGCGGCACTGGGTGACACCCACACTGGCCTGGCGCTGATGCGC 360
Db 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
QY 361 AAGGAACAGCTGTTGTTGAGACATCAGGTGCGCGCGCCAGGGGTGCCAAAGTGGTG 420
Db 121 LysGlnGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuVal 140
QY 421 TGGGTGACAGATGCGGCTCCAGCAGACCTGTGGGCCCCCGCATGAGAGCTCAAGAC 480
Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGlnLeuLysAsp 160
QY 481 CTGGGCGCTACCGTGTTCATTGTGACGACCGGCCAGAGCACTTCTTGAGAGCTGTACGC 540
Db 161 LeuGlyValThrValPheLeValSerThrGlyArgGlyAspPheLeuGlnLeuSerAla 180
QY 541 GCTGCTCAGCCCGCCGCGAGAGCACTGTTGAGAGTGGATGACCTGCACATC 600
Db 181 AlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspValAspPheLeuHisLe 200
QY 601 ATTGTCAAGAGCTGAGGGGCTCCATTCTC--GCGATGCGGCGCGAGCAGCTCCATGCC 657
Db 201 IleValGlnGlnLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
QY 658 ACGGAGATCAGCTCCAGCGGCTTCGCGCTGCGCTGGCCACCCCTGTGACCGCAAGCTCG 717
Db 221 ThrGlnIleThrSerSerGlyPheArgLeuAlaTrpProProLeuLeuThrAlaAspSer 240
QY 718 GCGTCACTATGCTGAGAGCTGAGTGCACGCGCCAGCGCGGGGCTGCAGAGACGACG 777
Db 241 GlyTyrTyrValLeuGlnLeuValProSerAlaGlnProGlyAlaAlaArgArgGlnGln 260
QY 778 CTGCCAGGAGAACGCGCAGCTGATCTGGCGCGGCTGCAGCCCGGACAGCACTACGAC 837
Db 261 LeuProGlyAsnAlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyrAsp 280
QY 838 GTGGCGCTAGTGCCTGAGTCAAAGTGCCTCTCTGAGGCCCCCAATCTCTCGGGTGC 897
Db 281 ValAlaLeuValProGlnSerAsnValArgLeuLeuArgProGlnIleLeuArgValArg 300
QY 898 ACGCGCGCAGAGAGCGCGCGCAGAGCGCATGTCATCTCCACAGCCCGCGCGCGAC 957
Db 301 ThrArgProGlnGlnAlaGlyProGlnArgIleValIleSerHisAlaArgProArgSer 320
QY 958 CTCCCGCTAGTGGGCGCCAGCGCTGAGCTGAGCCGCGCGCTCGGCTACACAGCGACG 1017
Db 321 LeuArgValSerTrpAlaProAlaLeuGlySerAlaAlaAlaLeuGlyTyrHisValGln 340
QY 1018 TTGGGCGCGCTGGGGGCGGGAGCGCAGCGGAGTGAAGTGCCTCGGGCGCGCACTGC 1077
Db 341 PheGlyProLeuArgGlyGlyGlnAlaGlnArgValGlnValProAlaGlyArgAsnCyS 360
QY 1078 ACCAAGCTGAGAGGCTCGCGCGCGCGCAGCGCTACCTGATGACCTGCACCGCGCTTC 1137
Db 361 ThrThrLeuGlnGlyLeuAlaProGlyThrAlaTyrLeuValThrValThrAlaAlaPhe 380

QY 1138 CGCTCGGCGCGCAGAGCGCGCTGTCCGCAAGGCTGCACGCGCCGAGCGCGCGCG 1197
Db 381 ArgSerGlyArgGlnSerAlaLeuSerAlaLysAlaCysThrProAspGlyProArgPro 400
QY 1198 CGCCCAAGCCCCGTTGCCCGCGCGCGCGCGCGCGCGGAGACCGCGCAGCGCTGACCG 1251
Db 401 ArgProArgProValProArgAlaProThrProGlyThrAlaSerValArgGlnPro 418
RESULT 2
US-10-699-035A-20
; Sequence 20, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838 0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PRA701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-20
Alignment Scores:
Pred. No.: 1,33e-115 Length: 418
Score: 2137.50 Matches: 417
Percent Similarity: 99.84 Mismatches: 0
Best Local Similarity: 99.88 Mismatches: 0
Query Match: 89.84 Gaps: 1
DB: 4
US-10-699-035A-5 (1-1254) X US-10-699-035A-20 (1-418)
QY 1 ATGCTCCCTGAGACGCGCGCTGGCGCTGAGCTTGGCGCTGGCGCGGAGC 60
Db 1 MetLeuProTPrThAlaLeuGlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
QY 61 GCGCGGAGCGCGCTCCACCAAGCATAGCCCCCGAGGGAGCTGATGTTCTCGTGAGAC 120
Db 21 GlyAlaGlnArgGlyProProAlaSerAlaProAlaGlyAspLeuMetPheLeuLeuAsp 40
QY 121 AGCTCAGGCGAGGCTCTCTACACTACAGATTCTCCGGGATTCCGGAGTTTGGGGGACGCTG 180
Db 41 SerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlnPheValGlyGlnLeu 60
QY 181 GTGGCTCACTGCCCCCTGGGGACCGGGGGCCCTGGCGCCAGCTGAGTGCATGCGGACGT 240
Db 61 ValAlaAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGCCATACACCGAGTTCCCTTCGCGCAGCAGCTCGGGTGAAGAGCTGCGCCAGGATGCG 300
Db 81 ArgProTyrThrGlnPheProPheGlyGlnHisSerSerGlyGlnAlaAlaGlnAspAla 100
QY 301 GTGGCTGCTTTCGCCAGCGGCACTGGGTGACACCCACACTGGCCTGGCGCTGATGCGC 360
Db 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
QY 361 AAGGAACAGCTGTTGTTGAGACATCAGGTGCGCGCGCCAGGGGTGCCAAAGTGGTG 420
Db 121 LysGlnGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuVal 140
QY 421 TGGGTGACAGATGCGGCTCCAGCAGACCTGTGGGCCCCCGCATGAGAGCTCAAGAC 480
Db 141 TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnGlnLeuLysAsp 160

```

QY 481 CTGGGCGTCAACCGTGTTCATTTGTCAGACCGGCGAGGCACTTCTGAGCTGTGACCC 540
DB 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 180
QY 541 GCTGCTCAACCGGCTGCGGAGGACCTGCACTTTGTGACGTGATGCTGACATC 600
DB 181 AlaAlaSerAlaProAlaGluValSerIleValAspPheValAspValAspPheLeuHisIle 200
QY 601 ATTGTCGAAGAGCTGAGGGGCTCCATTCTC--GCGATGCGGCGGAGAGCTCCATGCC 657
DB 201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgProGlnGlnLeuHisAla 220
QY 658 ACGAGATCAACGTCAACGCGGCTTCCGCTGCGCTGCGACCCCTGTCAGCCGAGACTCG 717
DB 221 ThrGluIleThrSerSerGlyPheArgLeuValIleTrpProLeuLeuThrAlaAspSer 240
QY 718 GGTACTATGTGTGTGAGCTGTGCTGCGGAGGCGGCGGAGGCGGCTGCAAGACCCAGAG 777
DB 241 GlyTyrTyrValLeuGluLeuValProSerAlaGlnProGlyAlaAlaArgArgGlnGln 260
QY 778 CTGCGCAGGGAACGCCAGCACTGATCTGGGCGGCGCTGACCCGAGACAGACTAGAC 837
DB 261 LeuProGlyAsnAlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyrAsp 280
QY 838 GTGGCGCTAGTGCCTGATCCAACTGACGCTGCGCTGCGGCGGCGGAGTCTGCGGCTGCGC 897
DB 281 ValAlaIleValAlaProGluSerAsnValArgLeuLeuArgProGlnIleLeuArgValArg 300
QY 898 ACGCGGCGCAGAGAGGCGCGGCGGCGGAGCGGAGCTGATCTGCTCCAGCGCGGCGGCGG 957
DB 301 ThrArgProGluGluAlaGlyProGluArgIleValIleSerHisAlaArgProArgSer 320
QY 958 CTCGCGGTGAGTGGGCGGCGGCGGCGGAGGCTGAGCGGCGGCGGCTGCGGCTGCAAGTGCAG 1017
DB 321 LeuArgValSerTrpAlaProAlaLeuGlySerAlaAlaLeuGlyTyrHisValGln 340
QY 1018 TTGCGGCGGCTGCGGCGGCGGAGGCGGAGCGGAGGAGGAGTGCCTGCGGCGGCGGCACTGC 1077
DB 341 PheGlyProLeuArgGlyGlyGluAlaGlnArgValGlnValProAlaGlyArgAsnGly 360
QY 1078 ACCAGCGTGAAGGCGGCTGCGGCGGCGGAGCGGCTGCTGATGCGGAGCGGCGCTTC 1137
DB 361 ThrThrLeuGlnGlnIleValAlaProGlyThrAlaTyrLeuValThrValThrAlaAlaPhe 380
QY 1138 CGGTGGGCGGCGGAGGCGGCTGCTGCGGAGGCGTGCAGCGGCGGCGGCGGCGGCGG 1197
DB 381 ArgSerGlyArgGluSerAlaLeuSerAlaValAlaCysThrProAspGlyProArgPro 400
QY 1198 CGGCCAGCGCCCGTGCCTGCGGCGGCGGAGCGGAGGCGGAGCGGCGGAGCGGCTGAGCGG 1251
DB 401 ArgProArgProValProAlaArgAlaProThrProGlyThrAlaSerArgGluPro 418

```

```

RESULT 3
US-10-699-035A-4
; Sequence 4, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838, 0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 415
; TYPE: PRT

```

```

; ORGANISM: Mus musculus
US-10-699-035A-4
Alignment Scores:
Pred. No.: 7,42e-87
Score: 1640.50
Percent Similarity: 85.9%
Best Local Similarity: 77.7%
Query Match: 68.9%
DB: 4 Gaps: 1
US-10-699-035A-5 (1-1254) x US-10-699-035A-4 (1-415)
QY 1 ATGCTCCCTGAGCGGCGCTCGGCTGAGCTTGCAGCTTGCAGCTGCGGCGGAGC 60
DB 1 MetLeuPheTrpThrAlaPheSerMetAlaLeuSerLeuArgGluAlaValArgSer 20
QY 61 GCGCGGAGCGGCGGCTCAGCAGTACAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 120
DB 21 SerIleGluArgGlySerThrAlaSerAspProGlnGlyAspLeuLeuPheLeuLeuAsp 40
QY 121 AGCTCAGCAGGCTCTCTCACTACAGAGTCTCCCGGAGTTGCGGAGTTGTGCGGAGCTG 180
DB 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCACTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGAGTCTGTCAGCTGAGCAGT 240
DB 61 ValAlaThrMetSerPheGlyProGlyAlaLeuArgAlaSerLeuValHisValGlySer 80
QY 241 CGGCGCTACACGAGTTCCTCCCTTGGCGGCGGAGCAGCAGTCCGAGGAGGAGGAGGAGTGC 300
DB 81 GlnProHisThrGluPheThrPheAspGlnTyrSerSerGlyGlnIleAlaIleArgAspAla 100
QY 301 GTCGCGGTCTTGCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
DB 101 IleArgValAlaProGlnArgMetGlyAspThrAsnThrGlyLeuAlaLeuAlaTyrAla 120
QY 361 AAGGAGCAGCTGTTTGTGCTGAAGATCAGGTCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 420
DB 121 LysGluGlnLeuPheAlaGluGluAlaGlyAlaArgProGlyValProLysValLeuVal 140
QY 421 TGGGTACAGATGAGGCGGCTTCCAGGAGCCTGTGAGGCGGCGGCGGAGGAGGAGGAGGAG 480
DB 141 TrpValThrAspGlyGlySerSerAspProValGlyProMetGlnGluLeuLysAsp 160
QY 481 CTGGGCGTCAACCGTGTTCATTTGTCAGACCGGCGGAGGAGCACTTCTGAGCTGTGACCC 540
DB 161 LeuGlyValThrIlePheIleValSerTrpGlyArgGlyAsnLeuLeuGluLeuAla 180
QY 541 GCTGCTCAACCGGCTGCGGAGGACCTGCACTTTGTGACGTGATGCTGACATC 600
DB 181 AlaAlaSerAlaProAlaGluValSerIleValAspPheValAspValAspPheLeuProIle 200
QY 601 ATTGTCGAAGAGCTGAGGGGCTCCATTCTC--GCGATGCGGCGGAGAGCTCCATGCC 657
DB 201 IleAlaArgGluLeuArgGlySerIleThrAspAlaMetGlnProGlnGlnLeuHisAla 220
QY 658 ACGAGATCAACGTCAACGCGGCTTCCGCTGCGCTGCGACCCCTGTCAGCCGAGACTCG 717
DB 221 SerGluValLeuSerSerGlyPheArgLeuSerTrpProLeuLeuThrAlaAspSer 240
QY 718 GGTACTATGTGTGTGAGCTGTGCTGCGGAGGCGGCGGAGGCGGCTGCAAGACCCAGAG 777
DB 241 GlyTyrTyrValLeuGluLeuValProSerAlaGlnProGlyAlaAlaArgArgGlnGln 260
QY 778 CTGCGCAGGGAACGCCAGCACTGATCTGGGCGGCGGAGGCGGAGGCGGAGTCTGAGCAG 837
DB 261 LeuProGlyAsnAlaThrAspTrpIleTrpAlaGlyLeuAspProAspThrAspTyrGlu 280
QY 838 GTGGCGCTAGTGCCTGATCCAACTGACGCTGCGGCGGCGGAGTCTGCGGCTGCGCAGTGC 897
DB 281 ValSerLeuLeuProGluSerAsnValHisLeuLeuArgProGlnHisValArgValArg 300

```

```
QY      898  ACGCGCCAGAGAGAGCCGGGCGCCAGAGCCATGTCATCTCCACGAGCCCGGCGGCGAGC  957
      |||
Db      301  ThrleuGlnGlnGlnAlaGlyProGlnuAArgIleValIleSerHisAlaArgProArgSer  320
QY      958  CTCGCCGTGAGTGGAGCCCGCCAGCGCTGAGCGCTGAGCGCGCGCTGCGGCTACACAGCTGACG  1017
      |||
Db      321  LeuArgValSerThrPheAlaProAlaLeuGlyProAlaSerHisLeuGlyTyrHisValGln  340
QY      1018  TTCGGGCGCGCTGCGGGGCGGGAGGCGAGCGGCTGAGGTGCGCGCGCGCGCAATGCG  1077
      |||
Db      341  LeuGlyProLeuGlnGlySerLeuGlnuAArgValGlnValProAlaGlyGlnIleuSer  360
QY      1078  ACCAGCGTGCAGAGGCTGCGCGCGCGCGCGCTACCTGCTGACCGCTGACCGCGCTTC  1137
      |||
Db      361  ThrThrValGlnGlyLeuThrProCysThrThrTyrLeuValThrValThrAlaAlaPhe  380
QY      1138  CGCTGCGGCGCGGAGAGCGCGCTGTCGCGCAAGAGCGCTGACCGCGCGCGCGCGCGCG  1197
      |||
Db      381  ArgSerGlyArgGlnArgAlaLeuSerAlaValAlaCysThrAlaSerGlyAlaArgThr  400
QY      1198  CGCCCAAGCCCGCTGCGCGCGCGCGCGCGAGCCCGGAGGACCGCGAGCGCG  1244
      |||
Db      401  Arg--AlaProGlnSerMetArgProGlnuAlaGlyProArgGlnuPro  415
```

RESULT 4

```
US-10-699-035A-21
; Sequence 21, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT US/A 071838, 0142
; CURRENT APPLICATION NUMBER: US/10/699, 035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU FR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-21
```

Alignment Scores:

```
Pred. No.: 7 42e-87 Length: 415
Score: 1640.50 Matches: 324
Percent Similarity: 85.9% Conservative: 34
Best Local Similarity: 77.7% Mismatches: 56
Query Match: 68.9% Indels: 3
DB: 4 Gaps: 1
```

US-10-699-035A-5 (1-1254) x US-10-699-035A-21 (1-415)

```
QY      1  ATGCTCCCTGAGACGCGCTCGGCTGAGCTTGCAGCTGCGCTGCGCGCGAGC  60
      |||
Db      1  MetLeuPheTrpThrAlaPheSerMetAlaLeuSerLeuArgLeuAlaLeuAlaArgSer  20
QY      61  GCGCGGAGCGCGGTCCACAGCATGAGCCCCCGAGGGGACCTGATGTTCTGCTGAC  120
      |||
Db      21  SerIleGluArgGlySerThrAlaSerAspProGlnGlyAspLeuIleuPheLeuIleuAsp  40
QY      121  AGCTACGCGAGCGTCTCTCACTACGAGTTCTCCCGGTTCCGGAGTTTGTGGGCGAGCTG  180
      |||
Db      41  SerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlnuPheValGlyGlnIleu  60
QY      181  GTGGCTCACTGCGCCCTGCGGAGCCGGGCGCTGCGCGAGCTGAGTGCAGCGGGCAGT  240
      |||
Db      61  ValAlaThrMetSerThrGlyProGlyAlaLeuArgAlaSerIleuValHisValGlySer  80
```

```
QY      241  CGGCGATACACCGAGTTCCCTTGGCCGACAGCAAGCTGGGTGAGGCTGCCAGATGCG  300
      |||
Db      81  GlnProHisThrGlnPheThrPheAspGlnTyrSerSerGlyGlnAlaIleArgAspAla  100
QY      301  GNGCGTGTCTTGCACCGCAGCATGGGTGACACCCACATGCGCTGCGCTGGTCTATGCG  360
      |||
Db      101  IleArgValAlaIleProGlnIleMetGlyAspThrAsnThrGlyLeuAlaIleuAlaTyrAla  120
QY      361  AAGGAACAGCTGTTTGTGAAGCATCAGGTGCGCGCGCGAGGGGTGCCAAAGTGTGTTG  420
      |||
Db      121  LysGlnGlnLeuPheAlaGlnGlnuAlaGlyAlaArgProGlyValProIleValLeuVal  140
QY      421  TGGGTCAGACATGGCGGCTCCAGGACCCCTGTGGGCGCGCGCGCGCGCGCGCGCGCG  480
      |||
Db      141  TrpValThrAspGlyGlySerSerAspProValGlyProPheMetGlnIleuIleuYAsp  160
QY      481  CTGGCGCTCACCGATTCATTGTCAGACCGCGCGAGGCACTTCTGTGAGTGTGACCG  540
      |||
Db      161  LeuGlyValIleThrIlePheIleValSerThrGlyArgGlyAsnLeuIleuGlnIleuAla  180
QY      541  GCTGCTCAGCCCTGCGGAGAGACCTGCACTTGTGACGTGATGACCTGCACTG  600
      |||
Db      181  AlaAlaSerAlaProAlaGlnuIleHisIleuHisPheValAspValAspAspLeuProIle  200
QY      601  ATTGTCAGAGCTGAGGGGCTCCATTCTC--GCGATGCGCGCGCGAGCAGCTCCATGCC  657
      |||
Db      201  IleAlaArgGlnLeuArgGlySerIleThrAspAlaMetGlnProGlnIleuHisAla  220
QY      658  ACGAGATCACGTCAGCGGCTTCCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG  717
      |||
Db      221  SerGlnuAlleuSerSerGlyPheArgLeuSerTrpProProIleuIleThrAlaAspSer  240
QY      718  GGTACTATGTGCTGGAAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG  777
      |||
Db      241  GlyTyrTyrValIleuGlnIleuValProSerGlyIleuAlaThrThrArgArgGlnGln  260
QY      778  CTGCCAGGGAACGCCAGCGATGATCTGGGCGCGCGCTCCAGCCGCGACAGCATACAGAC  837
      |||
Db      261  LeuProGlyAsnAlaThrSerTrpThrThrAspLeuAspProAspThrAspTyrGln  280
QY      838  GTGGCGTAGTGCCTGAGTCCAAAGTGCAGCTCCTGAGAGCCCGACAGTCTGCGGTGCGC  897
      |||
Db      281  ValSerLeuLeuProGlnSerHisValHisIleuLeuArgProGlnHisValArgValArg  300
QY      898  ACGCGCCAGAGAGAGCGCGGCGCGAGCGCATGTCATCTCCACGCGCGCGCGCGAGC  957
      |||
Db      301  ThrleuGlnGlnuAlaGlyProGlnuArgIleValIleSerHisAlaArgProArgSer  320
QY      958  CTCGCCGTGAGTGGAGCCCGCCAGCGCTGAGCGCTGAGCGCGCGCTGCGGCTACACAGT  1017
      |||
Db      321  LeuArgValSerThrPheAlaProAlaLeuGlyProAlaSerHisAlaLeuGlyTyrHisValGln  340
QY      1018  TTCGGGCGCGCTGCGGGGCGGGAGGCGAGCGGCTGAGGTGCGCGCGCGCGCAATGCG  1077
      |||
Db      341  LeuGlyProLeuGlnGlySerLeuGlnuArgValGlnValProAlaGlyGlnIleuSer  360
QY      1078  ACCAGCGTGCAGAGGCTGCGCGCGCGCGCGCTACCTGCTGACCGCTGACCGCGCTTC  1137
      |||
Db      361  ThrThrValGlnGlyLeuThrProCysThrThrTyrLeuValThrValThrAlaAlaPhe  380
QY      1138  CGCTGCGGCGCGGAGAGCGCGCTGTCGCGCAAGAGCGCTGACCGCGCGCGCGCGCGCG  1197
      |||
Db      381  ArgSerGlyArgGlnArgAlaLeuSerAlaValAlaCysThrAlaSerGlyAlaArgThr  400
QY      1198  CGCCCAAGCCCGCTGCGCGCGCGCGCGCGAGCCCGGAGGACCGCGAGCGCG  1244
      |||
Db      401  Arg--AlaProGlnSerMetArgProGlnuAlaGlyProArgGlnuPro  415
```

RESULT 5

```
US-09-789-561-85
; Sequence 85, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
```

APPLICANT: Ni et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: P2043p1
CURRENT APPLICATION NUMBER: US/09/789,561
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 85
LENGTH: 215
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (7)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-85

Alignment Scores:
Pred. No.: 1,03e-53 Length: 215
Score: 1067.50 Matches: 213
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 44.9% Indels: 1
DB: 3 Gaps: 1

US-10-699-035a-5 (1-1254) x US-09-789-561-85 (1-215)

QY 1 ATGCTCCCTGAGCGGCGCTGCGCTGAGCTTCCGCTGCGCGGAGC 60
DB 1 MetLeuProTyrThrAla**GlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
QY 61 GCGCGGAGCGCGGCTCACCAGCATCAGCCCCCGAGGGAGCCTGATGTTCTGCTGAC 120
DB 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
QY 121 AGCTAGCCAGCGCTCTCTCACTACGAGTTCTCCGGGTTCCGGAGTTTGTGGGGCAGCTG 180
DB 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACTGCGCTTGGGCAACCGGGGCTCGGTCAGTGTGTCAGCTGGGCACT 240
DB 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuAlaSerLeuValHisValGlySer 80
QY 241 CGGCAATACACGAGTTCCCTTCCGCGCAGCAGCTCGGGTGAAGCTGCCAGATGCG 300
DB 81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 100
QY 301 GTGGCTGCTTGGCCAGCGCATGGGTGACACCCACACTGCGCTGGCGTGTCTATGCC 360
DB 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
QY 361 AAGGAACAGCTGTTTCTGAAGCATCAGGTGCCCGGCGAGGGTGCCTCAAGTGTGGTG 420
DB 121 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuVal 140
QY 421 TGGGTGACAGATGGCGGCTCCAGGACCGCTGGGGCCCCCGCATGAGAGGCTCAAGGAC 480
DB 141 TrpValThrAspGlyLysSerSerAspProValGlyProProMetGlnGluLeuLysAsp 160
QY 481 CTGGGCGTCAACCGTGTTCATGTTCAGACCGCGCGAGCAACTCTCTGAGACTGTGACGC 540
DB 161 LeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPheLeuGluLeuSerAla 180
QY 541 GCTGCTCAGCCCTGCGGAGAGACACTGCACTTTGTGAGCTGGATGACCTGCAATC 600
DB 181 AlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspValAspAspLeuHisIle 200

QY 601 ATGTCCAGAGCTGAGGGGCTCCATTCTC--GCGATCGGCGC 642
DB 201 IleValGlnGluLeuArgGlySerIleLeuAspAlaMetArgPro 215

RESULT 6

US-09-833-245-2096
Sequence 2096, Application US/09833245
Publication No. US20040010134A1
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Albumin Fusion Proteins
FILE REFERENCE: PFS46PCT
CURRENT FILING DATE: US/09/833,245
PRIOR APPLICATION NUMBER: 60/229,358
PRIOR FILING DATE: 2000-04-12
PRIOR APPLICATION NUMBER: 60/256,931
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/199,384
PRIOR FILING DATE: 2000-04-25
NUMBER OF SEQ ID NOS: 2267
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2096
LENGTH: 215
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (7)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2096

Alignment Scores:

Pred. No.: 1,03e-53 Length: 215
Score: 1067.50 Matches: 213
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 44.9% Indels: 1
DB: 3 Gaps: 1

US-10-699-035a-5 (1-1254) x US-09-833-245-2096 (1-215)

QY 1 ATGCTCCCTGAGCGGCGCTGCGCTGAGCTTCCGCTGCGCGGAGC 60
DB 1 MetLeuProTyrThrAla**GlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 20
QY 61 GCGCGGAGCGCGGCTCACCAGCATCAGCCCCCGAGGGAGCCTGATGTTCTGCTGAC 120
DB 21 GlyAlaGluArgGlyProProAlaSerAlaProArgGlyAspLeuMetPheLeuLeuAsp 40
QY 121 AGCTAGCCAGCGCTCTCTCACTACGAGTTCTCCGGGTTCCGGAGTTTGTGGGGCAGCTG 180
DB 41 SerSerAlaSerValSerHisTyrGluPheSerArgValArgGluPheValGlyGlnLeu 60
QY 181 GTGGCTCCACTGCGCTTGGGCAACCGGGGCTCGGTCAGTGTGTCAGCTGGGCACT 240
DB 61 ValAlaProLeuProLeuGlyThrGlyAlaLeuAlaSerLeuValHisValGlySer 80
QY 241 CGGCAATACACGAGTTCCCTTCCGCGCAGCAGCTCGGGTGAAGCTGCCAGATGCG 300
DB 81 ArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGluAlaAlaGlnAspAla 100
QY 301 GTGGCTGCTTGGCCAGCGCATGGGTGACACCCACACTGCGCTGGCGTGTCTATGCC 360
DB 101 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 120
QY 361 AAGGAACAGCTGTTTCTGAAGCATCAGGTGCCCGGCGAGGGTGCCTCAAGTGTGGTG 420
DB 121 LysGluGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuVal 140
QY 421 TGGGTGACAGATGGCGGCTCCAGGACCGCTGGGGCCCCCGCATGAGAGGCTCAAGGAC 480
DB 141 TrpValThrAspGlyLysSerSerAspProValGlyProProMetGlnGluLeuLysAsp 160

QY 481 GTGGCGTCCAGCTGTTATTGTACAGACCGCCGAGCACTTCTGGAGCTGTACGC 540
Db 161 LeuGIyValThrValPheIleValSerThrGIyArgGIyAsnPhenIleuSerAla 180
QY 541 GCTGCTCAGCCCGCCGAGAGACCCCTGTCAGTGTGGAGTGTGACCTGACATC 600
Db 181 AlaAlaSerAlaProIleGluIleuShiSleuHisPheValAspValAspAspLeuHisIle 200
QY 601 ATTGTCCAAGAGCTGAGGGGCTCCATTCTC---GCCATGCGGCCG 642
Db 201 IleValGIingIleuArgGIySerIleIleuAspAlaMetArgPro 215

RESULT 7
US-10-883-936-85
; Sequence 85, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-85

Alignment Scores:
Pred. No.: 1,03e-53 Length: 215
Score: 1067.50 Matches: 213
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 44.9% Indels: 1
DB: 5 Gaps: 1

US-10-699-035A-5 (1-1254) x US-10-883-936-85 (1-215)

QY 1 ATGCTCCCTGGACGGCGCTCGGCTGCGCTGAGCTTGGCGCTGGCGCGGAGC 60
Db 1 MetIeuProTIPThAla**GIyeuAlaIeuSerIeuAlaArgIleuAlaIaArgSer 20
QY 61 GGCGGAGAGCGGCTCCACACGATCAGCCCGCCGAGGAGCCTGATGTCTCTGGAC 120
Db 21 GIyAlaGIuArgGIyProProAlaSerAlaProIleArgGIyAspIleuMetPheIleuAsp 40
QY 121 AGCTCAGGCGGCTCTCTGACTAGAGATTCTCCGGGATTGGGAGTTGTGGGAGCTG 180
Db 41 SerSerAlaSerValSerHisTyrGIuPheSerArgValArgIuPheValGIyGIuIleu 60
QY 181 GTGGCTCAGTGGCCCTGGGACCGGGGCTGCGCTGCTGTGTGACCTGGGCACT 240
Db 61 ValAlaProIleuPheIleuGIyThrGIyAlaIeuArgAlaSerIeuValHisValGIySer 80
QY 241 CGGCATACACGAGTTCCCTTGGCGACAGAGCTCGGTGAGGCTGCCAGAGATGCG 300
Db 81 ArgProTyrThrGIuPheProIleuGIyGIuHisSerSerGIyGIuAlaIaGIuIleuAla 100

QY 301 GTGGCTGCTTGTCCCGACCGCATGGGTGACACCCACATCGGCTGGCGCTTATGCC 360
Db 101 ValArgAlaSerAlaGlnArgMetGIyAspThrHisThrGIyIleuAlaIeuValTyrAla 120
QY 361 AAGAAAGAGCTGTTTGTGGAAGCATCAGTGGCCCGCCGAGGGGTGCCAAAGTGTGGTG 420
Db 121 LysGIuIleuPheAlaGIuAlaSerGIyAlaArgProGIyValProLysValIleuVal 140
QY 421 TGGGTGACAGATGGCGGCTCCAGCGACCTGTGGGCCCCCATGACAGAGCTCAGAGAC 480
Db 141 TrpValThrAspGIyGIySerSerAspProValGIyProPromeGIuIleuLysAsp 160
QY 481 CTGGCGTCCAGCCGRTTATTGTACAGACCGCGCCGAGCACTTCTGGAGCTGTACGC 540
Db 161 LeuGIyValThrValPheIleValSerThrGIyArgGIyAsnPhenIleuGIuIleuSerAla 180
QY 541 GCTGCTCAGCCCGCCGAGAGACCTGTCAGTGTGGAGTGTGACCTGACATC 600
Db 181 AlaAlaSerAlaProIleGluIleuShiSleuHisPheValAspValAspAspLeuHisIle 200
QY 601 ATTGTCCAAGAGCTGAGGGGCTCCATTCTC---GCCATGCGGCCG 642
Db 201 IleValGIingIleuArgGIySerIleIleuAspAlaMetArgPro 215

RESULT 8
US-09-789-561-159
; Sequence 159, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 159
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-159

Alignment Scores:
Pred. No.: 1,03e-53 Length: 242
Score: 1067.50 Matches: 213
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 44.9% Indels: 1
DB: 3 Gaps: 1

US-10-699-035A-5 (1-1254) x US-09-789-561-159 (1-242)

QY 1 ATGCTCCCTGGACGGCGCTCGGCTGCGCTGAGCTTGGCGCTGGCGCGGAGC 60
Db 28 MetIeuProTIPThAla**GIyeuAlaIeuSerIeuAlaArgIleuAlaIaArgSer 47
QY 61 GGCGGAGAGCGGCTCCACACGATCAGCCCGCCGAGGAGCCTGATGTCTCTGGAC 120

Db 48 GYALAGIARGLYProProAlaSerAlaProArgLYAspLeuMetPheLeuAsp 67
121 ACCTCAGCAGCGGTCTCTCACTACGAGTTCTCCCGGAGTTGGAGTTGGGGCAGCTG 180
QY 68 SerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlnPheValGlyGlnLeu 87
Db 181 GTGGCTCCACTGCGCCCTGGGCAACCGGGGCCCTGCGTGCAGTGTGTGACGTGGGCGAGT 240
QY 88 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 107
QY 241 CGGCATACACGAGTTCCCTTCCGCGCAGCAGTCTGGGAGGCTGCCAGATGGG 300
Db 108 ArgProTyrThrGlnPheProPheGlyGlnHisSerSerGlyGlnAlaGlnAspAla 127
QY 301 GTGGGCTCTTGGCCGAGCGCATGGGTGACACCCCACTGGCGCTGGCGCTGTATGGC 360
Db 128 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 147
QY 361 AAGGAACAGCTGTTTGTCTGAAGCATGAGTGGCCCGGCGCAAGGGGTGCCAAAGTGTGTG 420
Db 148 LysGlnGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyValProLysValLeuVal 167
QY 421 TGGGTACAGATGGCGGCTCTCCAGCAGCCTGTGGGGCCCCCATGACAGAGCTCAAGAC 480
Db 168 TrpValThrAspGlyGlySerSerAspProValGlyProMetGlnGlnLeuLysAsp 187
QY 481 CTGGGGGTACCGGTCTTCACTTGTGACAGCGGCGGCGCAACTTCTGGAGCTGTACGCC 540
Db 188 LeuGlyValThrValPheHisLeuAlaSerThrGlyArgGlyAsnPheLeuGlnLeuSerAla 207
QY 541 GGTGCTCAGCCCTGCGCGAGAGCACTGTGACCTTTGTGACGCTGATGACCTGCACATC 600
Db 208 AlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspValAspAspLeuHisIle 227
QY 601 ATTGTCCAAGAGCTGAGGGGCTTCATTCTC---GCCATGGCGCGG 642
Db 228 IleValGlnGlnLeuArgGlySerIleLeuAspAlaMetArgPro 242

RESULT 9
US-10-883-936-159
; Sequence 159, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: PZ043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; PRIOR FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 159
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE

LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-159
Alignment Scores:
pred. No.: 1,03e-53 Length: 242
Score: 1067.50 Matches: 213
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 44.9% Indels: 1
DB: 5 Gaps: 1

US-10-699-035a-5 (1-1254) x US-10-883-936-159 (1-242)
QY 1 ANGTCTCCCTGAGCAGCGGCTCGGCTGAGCTTGCAGCTTGCAGGCTGGCGCTGGCGCGAGC 60
Db 28 MetLeuProTyrPheAla***GlyLeuAlaLeuSerLeuArgLeuAlaLeuAlaArgSer 47
QY 61 GGGCGGAGCGCGGTCCACACGATCAGGCCCCCGAGGGAGCTGATGTTCTGCTGGAC 120
Db 48 GYALAGIARGLYProProAlaSerAlaProArgLYAspLeuMetPheLeuAsp 67
QY 121 AGCTCAGCAGCGGTCTCTCACTACGAGTTCTCCCGGAGTTGGAGTTGGGGCAGCTG 180
Db 68 SerSerAlaSerValSerHisTyrGlnPheSerArgValArgGlnPheValGlyGlnLeu 87
QY 181 GTGGCTCCACTGCGCCCTGGGCAACCGGGGCCCTGCGTGCAGTGTGTGACGTGGGCGAGT 240
Db 88 ValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerLeuValHisValGlySer 107
QY 241 CGGCATACACGAGTTCCCTTCCGCGCAGCAGTCTGGGAGGCTGCCAGATGGG 300
Db 108 ArgProTyrThrGlnPheProPheGlyGlnHisSerSerGlyGlnAlaGlnAspAla 127
QY 301 GTGGGCTCTTGGCCGAGCGCATGGGTGACACCCCACTGGCGCTGGCGCTGTATGGC 360
Db 128 ValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeuAlaLeuValTyrAla 147
QY 541 GGTGCTCAGCCCTGCGCGAGAGCACTGTGACCTTTGTGACGCTGATGACCTGCACATC 600
Db 208 AlaAlaSerAlaProAlaGlnLysHisLeuHisPheValAspValAspAspLeuHisIle 227
QY 601 ATTGTCCAAGAGCTGAGGGGCTTCATTCTC---GCCATGGCGCGG 642
Db 228 IleValGlnGlnLeuArgGlySerIleLeuAspAlaMetArgPro 242

RESULT 10
US-09-789-561-165
; Sequence 165, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: PZ043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315

```
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 165
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-789-561-165

Alignment Scores:
Pred. No.: 2,14e-53 Length: 226
Score: 1062.00 Matches: 210
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 44.6% Indels: 0
DB: 3 Gaps: 0

US-10-699-035A-5 (1-1254) x US-09-789-561-165 (1-226)

QY 1 ATGCTCCCTGGAGCGGCGCTCGGCTGAGCTTGCGGCTGCGGCGGAGC 60
   |||
Db 14 MetLeuProTPrThAlaLeuGlYleuAlaLeuSerLeuAlaLeuAlaArgSer 33
   |||
QY 61 GCGGCGAGCGCGGTTCACACAGATCAGCCCGGAGGAGCTGATGTTCTGTCGAC 120
   |||
Db 34 GlYAlaGlUArgGlYProProAlaSerAlaProAlaArgLYAspLeuMetPheLeuLeuAsp 53
   |||
QY 121 AGCTAGCGCAGCGCTCTCCTACAGAGTTCTCCCGGATTCCGGAGTTTGAGGAGCAGCTG 180
   |||
Db 54 SerSerAlaSerValSerHisTyrGlUupheSerAlaArgGlUupheValGlYleuLeu 73
   |||
QY 181 GTGGCTCCACTGCCCCCTGGGCAACCGGAGCCCTGCGTGCAGTCTGAGTGCAGTGGGCACT 240
   |||
Db 74 ValAlaProLeuProLeuGlYThrGlYAlaLeuArgAlaSerLeuValHisValGlYSer 93
   |||
QY 241 CGGCGCATACACCGAGTTCCCTTCCGCGCAGCAGCTCGGAGGAGCTGCCAGAGATGG 300
   |||
Db 94 ArgProLYrThnGluPheProPheGlYlnHisSerSerGlYlnAlaAlaGlInAspAla 113
   |||
QY 301 GTGCGTGTCTTGCAGCGCAGCATGGGTGACACCCACACTGCGCTGGCGCTGCTATGCC 360
   |||
Db 114 ValAlaGlaSerAlaGlInArgMetGlYAspThrHisThrGlYleuAlaLeuValTYrAla 133
   |||
QY 361 AAGAACAGCTGTTTGTCTGAAGCATCAGTGCCCGGCGCAGGGGTGCCAAAGTGTGCTG 420
   |||
Db 134 LysGlUlnLeuPheAlaGlInAlaSerGlYAlaArgProGlYValProLYsValLeuVal 153
   |||
QY 421 TGGGTGACAGATGGGCGCTCCAGCAGCCTGTGGGCGCCCGCATGAGAGAGCTCAAGGAC 480
   |||
Db 154 TrpValThrAspGlYGlYSerSerAspProValGlYProPheMetGlInLeuLeuYAsp 173
   |||
QY 481 CTGGCGCTGACCGCTGTTCTGATTCAGACACCGGCGCAGGCAACTTCTTGAGGCTGCAGCC 540
   |||
Db 174 LeuGlYValThrValPheIleValSerThrGlYArgGlYAsnPheLeuGlInLeuSerAla 193
   |||
QY 541 GCTGCTCAGCCCTTCGCGAGAGACACTTGTGAGCGTGGATGACCTGCACATC 600
   |||
Db 194 AlaAlaSerAlaProAlaGlUlyshisLeuHisPheValAspValAspAspLeuHisIle 213
   |||
QY 601 ATTGTCCAAGAGCTGAGGGGCTCCATTCTC 630
   |||
Db 214 IleValGlInGluLeuArgGlYSerIleLeu 223
   |||

RESULT 11
US-10-883-936-165
; Sequence 165, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043PI
; CURRENT APPLICATION NUMBER: US/10/883,936
; CURRENT FILING DATE: 2004-07-06
```

```
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 165
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-883-936-165

Alignment Scores:
Pred. No.: 2,14e-53 Length: 226
Score: 1062.00 Matches: 210
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 44.6% Indels: 0
DB: 5 Gaps: 0

US-10-699-035A-5 (1-1254) x US-10-883-936-165 (1-226)

QY 1 ATGCTCCCTGGAGCGGCGCTCGGCTGAGCTTGCGGCTGCGGCGGAGC 60
   |||
Db 14 MetLeuProTPrThAlaLeuGlYleuAlaLeuSerLeuAlaLeuAlaArgSer 33
   |||
QY 61 GCGGCGAGCGCGGTTCACACAGATCAGCCCGGAGGAGCTGATGTTCTGTCGAC 120
   |||
Db 34 GlYAlaGlUArgGlYProProAlaSerAlaProAlaArgLYAspLeuMetPheLeuLeuAsp 53
   |||
QY 121 AGCTAGCGCAGCGCTCTCCTACAGAGTTCTCCCGGATTCCGGAGTTTGAGGAGCAGCTG 180
   |||
Db 54 SerSerAlaSerValSerHisTyrGlUupheSerAlaArgGlUupheValGlYleuLeu 73
   |||
QY 181 GTGGCTCCACTGCCCCCTGGGCAACCGGAGCCCTGCGTGCAGTCTGAGTGCAGTGGGCACT 240
   |||
Db 74 ValAlaProLeuProLeuGlYThrGlYAlaLeuArgAlaSerLeuValHisValGlYSer 93
   |||
QY 241 CGGCGCATACACCGAGTTCCCTTCCGCGCAGCAGCTCGGAGGAGCTGCCAGAGATCG 300
   |||
Db 94 ArgProLYrThnGluPheProPheGlYlnHisSerSerGlYlnAlaAlaGlInAspAla 113
   |||
QY 301 GTGCGTGTCTTGCAGCGCAGCATGGGTGACACCCACACTGCGCTGGCGCTGCTATGCC 360
   |||
Db 114 ValAlaGlaSerAlaGlInArgMetGlYAspThrHisThrGlYleuAlaLeuValTYrAla 133
   |||
QY 361 AAGAACAGCTGTTTGTCTGAAGCATCAGTGCCCGGCGCAGGGGTGCCAAAGTGTGCTG 420
   |||
Db 134 LysGlUlnLeuPheAlaGlInAlaSerGlYAlaArgProGlYValProLYsValLeuVal 153
   |||
QY 421 TGGGTGACAGATGGGCGCTCCAGCAGCCTGTGGGCGCCCGCATGAGAGAGCTCAAGGAC 480
   |||
Db 154 TrpValThrAspGlYGlYSerSerAspProValGlYProPheMetGlInLeuLeuYAsp 173
   |||
QY 481 CTGGCGCTGACCGCTGTTCTGATTCAGACACCGGCGCAGGCAACTTCTTGAGGCTGCAGCC 540
   |||
Db 174 LeuGlYValThrValPheIleValSerThrGlYArgGlYAsnPheLeuGlInLeuSerAla 193
   |||
QY 541 GCTGCTCAGCCCTTCGCGAGAGACACTTGTGAGCGTGGATGACCTGCACATC 600
   |||
Db 194 AlaAlaSerAlaProAlaGlUlyshisLeuHisPheValAspValAspAspLeuHisIle 213
   |||
QY 601 ATTGTCCAAGAGCTGAGGGGCTCCATTCTC 630
   |||
Db 214 IleValGlInGluLeuArgGlYSerIleLeu 223
   |||

RESULT 12
US-09-789-561-160
; Sequence 160, Application US/09789561
```

```
; Patent No. US2002006481BA1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043PI
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 160
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (152)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (180)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (184)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (186)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-160

Alignment Scores:
Pred. No.: 1,73e-44 Length: 186
Score: 908.00 Matches: 156
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 38.2% Indels: 0
DB: Gaps: 0

US-10-699-035A-5 (1-1254) x US-09-789-561-160 (1-186)
QY 472 GCTCTGATGGGGGGGGCCACAGGATCGCTGAGCGCATGTGACCCACAGCA 413
DB 1 AlaProAlaTrpGlyGlnGlyProGlnGlyArgTrpSerArgHisLeuSerProThrProAla 20
QY 412 CTTTGGGACCCCTGGCCGGGACCTGATGCTTCAGCAAAACAGCTGTTCTTGGCATAGA 353
DB 21 LeuTrpAlaProLeuAlaGlnHisLeuMetLeuGlnGlnThrAlaValProThrPheArg 40
QY 352 CCAGCGCCAGGCGGAGTGGGTGTCAACCATGCGCTGGGGGAGAGCAAGCAGCGCATCCT 293
DB 41 ProAlaProGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 60
QY 292 GGGCAGCTCAGCCGAGCTGTGCTGGCCGAGGAGGAACTGATGTTATGGCCGAGCTGCCA 233
DB 61 GlyGlnProHisProSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 80
QY 232 CGTGACCAAGACTGGACGAGGAGGCGCGGTGGCCAGGGGACAGTGAAGCCACAGCTGCC 173
DB 81 ArgAlaProAspTrpHisAlaGlnGlyProArgGlyProGlyAlaValGlnProProAlaAla 100
QY 172 CCAGAACTCCGAGACCCGGGAGAACTGTAAGTGAAGAGAGCTGGCTGAGCTGTCCAGCA 113
DB 101 ProGlnThrProGlnProGlnArgTrpArgSerGlnArgTrpLeuSerGlyProAla 120
QY 112 GGAACATCAGGCTCCCTGGGGGGGCTGATGCTGTGAGACCGGCTCCGCGCGCTCCGCG 53
DB 121 GlyThrSerGlyProLeuGlnGlyGlyLeuMetLeuValAlaPheArgAlaProHisArgSerAla 140
QY 52 CCAGCGCCAGCGCCAGAGCTCAGGGCCAGGCCGAGCGCGCTCAGGGGAGCA 2
```

```
DB 141 ProAlaProAlaAlaSerSerGlyProGlyArg***ProSerArgGlyAla 157

RESULT 13
US-10-883-936-160
; Sequence 160, Application US/10883936
; Publication No. US2005001986A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043PI
; CURRENT APPLICATION NUMBER: US/10/883,936
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 160
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (152)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (180)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (184)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (186)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-160

Alignment Scores:
Pred. No.: 1,73e-44 Length: 186
Score: 908.00 Matches: 156
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 38.2% Indels: 0
DB: Gaps: 0

US-10-699-035A-5 (1-1254) x US-10-883-936-160 (1-186)
QY 472 GCTCTGATGGGGGGGGCCACAGGATCGCTGAGCGCATGTGACCCACAGCA 413
DB 1 AlaProAlaTrpGlyGlnGlyProGlnGlyArgTrpSerArgHisLeuSerProThrProAla 20
QY 412 CTTTGGGACCCCTGGCCGGGACCTGATGCTTCAGCAAAACAGCTGTTCTTGGCATAGA 353
DB 21 LeuTrpAlaProLeuAlaGlnHisLeuMetLeuGlnGlnThrAlaValProThrPheArg 40
QY 352 CCAGCGCCAGGCGGAGTGGGTGTCAACCATGCGCTGGGGGAGAGCAAGCAGCGCATCCT 293
DB 41 ProAlaProGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 60
QY 292 GGGCAGCTCAGCCGAGCTGTGCTGGCCGAGGAGGAACTGATGTTATGGCCGAGCTGCCA 233
DB 61 GlyGlnProHisProSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 80
QY 232 CGTGACCAAGACTGGACGAGGAGGCGCGGTGGCCAGGGGACAGTGAAGCCACAGCTGCC 173
```

```
Db      81 ArgAlaProAbpTrpHisAlaGlyProAbgCysProGlyAlaValGluProProAlaAla 100
Qy      172 CCACAACCTCCCGAAGCCGGGAGAACTGTAAGAGACGCTGGCTGAGCTGTCCAGCA 113
Db      101 ProGlnThrProGluProGlyArgThrArgSerGluArgArgTrpLeuSerCysProAla 120
Qy      112 GGAACATCAAGTCCCTCGGGGGGCTGATGCTGTGTGACCGCGCTCCGGCGCTCCGG 53
Db      121 GlyThrSerGlyProLeuGlyGlyLeuMetLeuValAspArgAlaProArgArgSerAla 140
Qy      52 CCAGGCGCAGCCGCAAGCTCAGGGCCGAGGGCGGCTCCAGGGGAGCA 2
Db      141 ProAlaProAlaAlaSerSerCysProGlyArg***ProSerArgGlyAla 157

RESULT 14
US-10-699-035A-2
; Sequence 2, Application US/10699035A
; Publication No. US2004021439A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-2

Alignment Scores:
Pred. No.: 1,97e-44 Length: 180
Score: 907.00 Matches: 179
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 38.1% Indels: 0
DB: 4 Gaps: 0

US-10-699-035A-5 (1-1254) x US-10-699-035A-2 (1-180)
Qy      94 CGAGGGAGACTATGTTCTGCTGGACAGCTCAGCCAGGCTCTTCACATCAGAGTTTCTCC 153
Db      1 ArgGlyAspLeuMetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSer 20
Qy      154 CGGGTTCCGGAGTTTGTGGGAGCTGGTGGCTCCACTGCCCTCCGAGGAGCCGGGCGCTG 213
Db      21 ArgValAlaArgGluPheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeu 40
Qy      214 CGTGCAGTCTGTGTGACGCTGGGAGTGGCCATACCCGAGTCTCCCTTCGGCCAGCAC 273
Db      41 ArgAlaSerLeuValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHis 60
Qy      274 AGCTGGGGTGAAGGCTGCCAGAGATGGGTCGCTTCGCCAGGCGATGGGTGACACC 333
Db      61 SerSerGlyGlnAlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThr 80
Qy      334 CACACTGGCGCTGGCGCTGTATGSCCAAGAAAGCTTTGTCTGAACATCAGGTGCC 393
Db      81 HisThrGlyLeuAlaLeuValTyrAlaGlyGlnLeuPheAlaGlnAlaSerGlyAla 100
Qy      394 CGGCGAGGGGTGCCAAAGTGTGTGTGGTGAACAGATGGCGGCTCCAGCGACCTGTG 453
Db      101 ArgProGlyValProGlyValLeuValTyrValThrAspGlyGlySerSerAspProVal 120
Qy      454 GGGCCCCCATGACAGAGCTCAAGAGCTGGGGCGTACCGTGTCTATTGTACGACCGGC 513
```

```
Db      121 GlyProPheMetGlnGluLeuLeuAspLeuGlyValThrValPheIleValSerThrGly 140
Qy      514 CGAGGCAACTTCTCTGAGCTGTCAAGCCGCTGCTCAGCCCTGCGCGAGAACACCTGCAC 573
Db      141 ArgGlyAsnPheLeuGlnLeuSerAlaAlaAlaSerAlaProAlaGlnGlyHisIleuHis 160
Qy      574 TTTGTGACGTGATGACCTGCACATCATTTGTCCAGAAGCTGAGGGGCTCCATTCTC 630
Db      161 PheValAspValAlaAspAspLeuHisIleIleValGlnGluLeuArgGlySerIleu 179

RESULT 15
US-10-149-819-18
; Sequence 18, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyoung Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAU, Preci
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 18
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3683905CD1
US-10-149-819-18

Alignment Scores:
Pred. No.: 4.1e-44 Length: 185
Score: 901.50 Matches: 180
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 0
Query Match: 37.9% Indels: 1
DB: 4 Gaps: 1

US-10-699-035A-5 (1-1254) x US-10-149-819-18 (1-185)
Qy      106 ATGTTCCGCTGGAGACAGTCAAGCCAGGCTCTCTCACTCAAGATTCTCCGGGTTCCGGAG 165
Db      1 MetPheLeuLeuAspSerSerAlaSerValSerHisTyrGluPheSerArgValArgGlu 20
Qy      166 TTTGTGGGAGCACTGTGGTCTCACTGCCCTTGAGGACCGGGGCGCTCGTGCACAGTCTG 225
Db      21 PheValGlyGlnLeuValAlaProLeuProLeuGlyThrGlyAlaLeuArgAlaSerIleu 40
Qy      226 GTGCAGTGGGCAGTGGCCATPACCCGAGTTTCCCTTCGGCCAGACACAGCTCGGGTCA 285
Db      41 ValHisValGlySerArgProTyrThrGluPheProPheGlyGlnHisSerSerGlyGlu 60
Qy      286 GCTGCCAGATGCGGTCGCTTCGCTTCGCCAGGCGAGGGTGAACCCAGACCTGGCGCTG 345
Db      61 AlaAlaGlnAspAlaValArgAlaSerAlaGlnArgMetGlyAspThrHisThrGlyLeu 80
Qy      346 GCGCTGCTCTATGCAAGAAAGAGCTTTGTCTGAACATCAGGTGCCCGGCGAGGGGTG 405
Db      81 AlaLeuValTyrAlaGlyGlnGlnLeuPheAlaGlnAlaSerGlyAlaArgProGlyVal 100
```

Qy	406	CCCAAGTCGTGCTGGGTGACAGATGGCGGCTCCAGCGACCCCTGTGGGCCCCCATG	465
Db	101	ProlyValIleuValItrpValThrAspGlyGlySerSerAspProValGlyProPomMet	120
Qy	466	CAGAGCTCAAGAGCCTGGGCGGTCAACGTTGATTCAGACACCGCGAGGCACTTC	525
Db	121	GInGluLeuLysAspLeuGlyValThrValPheIleValSerThrGlyArgGlyAsnPro	140
Qy	526	CTGAGCTGTCAAGCCGCTGCTCAGCCCTGCCGAGAGACACCTGCACCTTTGTGACGTG	585
Db	141	LeuGluLeuSerAlaIleAlaIleSerAlaProAlaGluLysHisLeuHisPheValAspVal	160
Qy	586	GATGACCTGCACATCATTTGTCCAGAGCTGAGGGGCTCCATTCTC--GCGATGGGCGCG	642
Db	161	AspAspLeuHisIleIleValGInGluLeuArgGlySerIleLeuAspAlaMetArgPro	180
Qy	643	CAG 645	
Db	181	Gln 181	

Search completed: February 13, 2006, 13:53:31
 Job time : 212.799 secs

THIS PAGE BLANK (USPTO)

CC molecular marker, used for detecting a loss of ECM integrity in an animal)

CC subject, monitoring repair, regeneration or other disease processes in an
 CC animal subject and detecting a disease condition or a propensity for the
 CC development of a disease condition in an animal subject. The invention is
 CC useful for the manufacture of a medicament in the treatment of a disease
 CC condition of the ECM. The disease condition involves the cartilage, and
 CC is preferably arthritis. The invention is also used in gene therapy. The
 CC present sequence is human VA domain

XX Sequence 180 AA;

Query Match 100.0%; Score 913; DB 6; Length 180;
 Best Local Similarity 100.0%; Pred. No. 1.6e-94;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREVGQVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 60
 DB 1 RGDLMFLDSSASVSHYEFSSRVREVGQVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 60
 QY 61 SSGEAAQDAVAVASAGRMGDTHGTALVYAKQQLFAEASGARGPVKVLWVWDGSSDPV 120
 DB 61 SSGEAAQDAVAVASAGRMGDTHGTALVYAKQQLFAEASGARGPVKVLWVWDGSSDPV 120
 QY 121 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSTLD 180
 DB 121 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSTLD 180

RESULT 2

AAAB87344 ID AAAB87344 standard; protein; 215 AA.

XX AAAB87344;

DT 22-MAY-2001 (first entry)

DE Human gene 3 encoded secreted protein HMTB078, SEQ ID NO:85.

XX Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; viderary; cell culture;
 KW chemotaxis; food additive; binding partner identification.

XX Homo sapiens.

XX WO200118022-A1.

XX 15-MAR-2001.

XX 31-AUG-2000; 2000WO-US024008.

XX 03-SEP-1999; 99US-0152315P.

XX 03-SEP-1999; 99US-0152317P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Baker KP, Birse CE, Piscella M, Komatsoulis GA, Rosen CA;
 PI Sopet DR, Young PE, Ebnner R, Duan DR, Olsen HS, Lafleur DW;
 PI Moore PA, Shi Y, Wei Y, Florence KA;

XX WPI; 2001-203081/20.

XX N-PSDB; AAF91860.

PT Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers.

PS Claim 11; Page 532-533; 607pp; English.

XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
 CC protein genes, and AAAB87342-AAAB7413 represent the proteins they encode.
 CC AAAB87414-AAAB7454 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 52 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
 CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention

XX Sequence 215 AA;

Query Match 100.0%; Score 913; DB 4; Length 215;
 Best Local Similarity 100.0%; Pred. No. 2e-94;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREVGQVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 60
 DB 32 RGDLMFLDSSASVSHYEFSSRVREVGQVAPLPGLGTALRASLVHVGSRPYTEFPFGQH 91
 QY 61 SSGEAAQDAVAVASAGRMGDTHGTALVYAKQQLFAEASGARGPVKVLWVWDGSSDPV 120
 DB 92 SSGEAAQDAVAVASAGRMGDTHGTALVYAKQQLFAEASGARGPVKVLWVWDGSSDPV 151
 QY 121 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSTLD 180
 DB 152 GPPMOELKDLGVTVFTVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSTLD 211

RESULT 3

ABG65347 ID ABG65347 standard; protein; 215 AA.

XX ABG65347;

DT 27-AUG-2002 (first entry)

DE Human albumin fusion protein #2022.

XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective tissue;
 KW cytoskeletal; antifertility; antiinflammatory; anticancer;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.

XX Homo sapiens.
 OS Synthetic.

XX WO200177137-A1.

XX 18-OCT-2001.
PD
XX
XX 12-APR-2001; 2001WO-US011988.
PF
XX
XX 12-APR-2000; 2000US-0229358P.
PR
XX 25-APR-2000; 2000US-0199384P.
PR
XX 21-DEC-2000; 2000US-0256931P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Haseltine WA;
XX
XX WPI; 2002-010886/01.
DR
XX
XX New fusion protein for treating disease e.g. diabetes comprises an
PT albumin fused to a therapeutic protein.
PS
XX Claim 1; Page 1935; 2102pp; English.
XX
XX The present invention relates to albumin fusion proteins comprising a
CC therapeutic protein X and human albumin (HA, also known as human serum
CC albumin, HSA). The proteins are useful for treating a disease or disorder
CC that may be modulated by therapeutic protein X. The albumin extends the
CC shelf-life of protein X, and may increase its biological in vitro/in vivo
CC activity. The protein is useful for treating and diagnosing disorders
CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
CC disease, ulcerative colitis), immune disorders (e.g. acquired
CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,
CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
CC schizophrenia), and connective tissue disorders (e.g. osteoporosis, arthritis).
CC ABG63326-ABG65518 represent albumin fusion proteins of the invention
XX
SQ Sequence 215 AA;

Query Match 100.0%; Score 913; DB 5; Length 215;
Best Local Similarity 100.0%; Pred. No. 2e-94;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RDDMFLDSSASVSHYERSRREPFGOLVAPLPCTGTLRLSLVWVGSRPTEPPFGQH 60
DB 32 RDDMFLDSSASVSHYERSRREPFGOLVAPLPCTGTLRLSLVWVGSRPTEPPFGQH 91
QY 61 SSGEAAQDVRASAOQMGDTHGLAVYAKEDLPFAASGAPGVPKVLVWVDGSSDPV 120
DB 92 SSGEAAQDVRASAOQMGDTHGLAVYAKEDLPFAASGAPGVPKVLVWVDGSSDPV 151
QY 121 GPPMELKDLGYTVFIVSTGRGNFLELSAASAPAEKHLFPVDVDLHIIVOELRSGIID 180
DB 152 GPPMELKDLGYTVFIVSTGRGNFLELSAASAPAEKHLFPVDVDLHIIVOELRSGIID 211

RESULT 4
ADL78614 ID ADL78614 strand; protein; 215 AA.
XX
XX ADL78614;
XX
XX 20-MAY-2004 (first entry)
XX
XX Albumin fusion protein related therapeutic protein X, SEQ ID No 2096.
XX
XX albumin fusion protein; cytostatic; antianaemic; antiarthritic;
XX antiaesthetic; anti-HIV; immunosuppressive; antiinflammatory;
XX antiparasitic; antibacterial; osteopathic; dermatological; antitumor;
XX immunomodulator; antiarrhythmic; cardiac; neurotropic; antidiabetic;
XX nephrotoxic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
XX antidiabetic; anabolic; hypertensive; vulnery; gene therapy; cancer;
XX reproductive system disorder; therapeutic protein.
XX
XX Unidentified.
XX

PN US2004010134-A1.
XX
XX 15-JAN-2004.
PD
XX
XX 12-APR-2001; 2001US-00833245.
PF
XX
XX 12-APR-2000; 2000US-0229358P.
PR
XX 25-APR-2000; 2000US-0199384P.
PR
XX 21-DEC-2000; 2000US-0256931P.
XX
XX (ROSE/) ROSEN C A.
PA (HASE/) HASELTINE W A.
XX
XX Rosen CA, Haseltine WA;
XX
XX WPI; 2004-090519/09.
DR
XX
XX New albumin fusion proteins, useful for diagnosing, treating, preventing
PT or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,
PT asthma, inflammatory bowel disease or Alzheimer's disease.
PT
XX
XX Disclosure; SEQ ID NO 2096; 279pp; English.
XX
XX The invention relates to a novel albumin fusion protein. The invention
CC further relates to: a composition comprising the albumin fusion protein
CC and a pharmaceutical carrier; a kit comprising the composition of the
CC albumin fusion protein formula; a method of treating a disease or
CC disorder in a patient comprising the step of administering the albumin
CC fusion protein; a method of treating a patient with a disease or disorder
CC that is modulated by therapeutic protein X, or its fragment or variant;
CC a method of extending the shelf life of therapeutic protein X, or its
CC fragment or variant; a nucleic acid molecule comprising a polynucleotide
CC sequence encoding the albumin fusion protein; a vector comprising the
CC nucleic acid molecule of the albumin fusion protein; and a host cell
CC comprising the nucleic acid molecule of the albumin fusion protein. The
CC albumin fusion protein and its compositions have the following
CC activities: cytostatic, antianaemic, antiarthritic, antiaesthetic, anti-
CC HIV, immunosuppressive, antiinflammatory, antiparasitic, antibacterial,
CC osteopathic, dermatological, antitumor, immunomodulator, antiarrhythmic,
CC cardiac, neurotropic, antidiabetic, nephrotoxic, uropathic,
CC neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic,
CC hypertensive, and vulnery. The albumin fusion protein nucleic acid may
CC be used in gene therapy to treat disorders. The albumin fusion protein is
CC useful for diagnosing, treating, preventing or ameliorating diseases or
CC disorders comprising indication: Y. The diseases or disorders include:
CC cancer (e.g. leukaemia, colon, bone, breast, liver or lung cancer),
CC immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute
CC lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS),
CC autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
CC disease), reproductive system disorders (e.g. prostatitis, inguinal
CC hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
CC leydig tumours), musculoskeletal diseases (e.g. giant cell tumours,
CC Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy
CC or cachexia), cardiovascular disease (e.g. Thrombomys, heart disease,
CC arrhythmia, cardiac arrest, heat valve disease, hypernatremia or
CC hyponatremia), mixed foetal diseases (e.g. foetal alcohol syndrome,
CC Down's syndrome, Parau syndrome, Turner's syndrome, Apert syndrome or Tay-
CC Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
CC tract infections or renal disorders), neural or sensory disease (e.g.
CC Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
CC cerebellar ataxia, attention deficit disorder, autism or obsessive
CC compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or
CC occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
CC disease or glomerulonephritis), digestive diseases (e.g. portal
CC hypertension, irritable bowel disease, gastric atrophy or pancreatitis)
CC or connective tissue or epithelial diseases (e.g. Crohn's disease,
CC scleroderma, wound healing or epidermolysis bullosa). This sequence
CC represents a therapeutic protein X relating to the albumin fusion protein
CC of the invention. The sequence listing data for this specification was
CC downloaded from the USPTO website.
XX
XX Sequence 215 AA;
SQ

Query Match 100.0%; Score 913; DB 8; Length 215;
Best Local Similarity 100.0%; Pred. No. 2e-94;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDLMFLDSSASVSHYEFSPRREFVQGVLPPLGTGALRASLVHVSRRPYTFPPFQOH 60
Db 32 RGDLMFLDSSASVSHYEFSPRREFVQGVLPPLGTGALRASLVHVSRRPYTFPPFQOH 91
Qy 61 SSGEAAQDAVBAQAQRMGDTHTGTALVYAKQQLFAEASGARGVPKVLVWTTDGGSSDPV 120
Db 92 SSGEAAQDAVBAQAQRMGDTHTGTALVYAKQQLFAEASGARGVPKVLVWTTDGGSSDPV 151
Qy 121 GPPMOELKDLGVTVPIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
Db 152 GPPMOELKDLGVTVPIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211

RESULT 5
AAB87424
ID AAB87424 standard; protein; 226 AA.
AC AAB87424;
XX
XX
XX 22-MAY-2001 (first entry)
XX
XX Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vunerary; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
XX Homo sapiens.
OS
XX
XX WO200118022-A1.
PN
XX
XX 15-MAR-2001.
PD
XX
XX 31-AUG-2000; 2000WO-US024008.
PF
XX
XX 03-SEP-1999; 99US-0152315P.
PR
XX
XX 03-SEP-1999; 99US-0152317P.
XX
XX (HUMA-) HUMAN GENOME SCT INC.
PA
XX
XX Ni J, Baker KP, Birse CE, Piscella M, Komatsoulis GA, Rosen CA,
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW,
PI Moore PA, Shi Y, Wei Y, Florence KA;
PI
XX WPI; 2001-203081/20.
DR
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
PT
XX
XX Disclosure; Page 18; 607P; English.
PS
XX
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AA87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of

CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX SQ Sequence 226 AA;

Query Match 100.0%; Score 913; DB 4; Length 226;
Best Local Similarity 100.0%; Pred. No. 2.2e-94;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDLMFLDSSASVSHYEFSPRREFVQGVLPPLGTGALRASLVHVSRRPYTFPPFQOH 60
Db 45 RGDLMFLDSSASVSHYEFSPRREFVQGVLPPLGTGALRASLVHVSRRPYTFPPFQOH 104
Qy 61 SSGEAAQDAVBAQAQRMGDTHTGTALVYAKQQLFAEASGARGVPKVLVWTTDGGSSDPV 120
Db 105 SSGEAAQDAVBAQAQRMGDTHTGTALVYAKQQLFAEASGARGVPKVLVWTTDGGSSDPV 164
Qy 121 GPPMOELKDLGVTVPIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
Db 165 GPPMOELKDLGVTVPIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 224

RESULT 6
AAB87418
ID AAB87418 standard; protein; 242 AA.
AC AAB87418;
XX
XX
XX 22-MAY-2001 (first entry)
XX
XX Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vunerary; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
XX Homo sapiens.
OS
XX
XX WO200118022-A1.
PN
XX
XX 15-MAR-2001.
PD
XX
XX 31-AUG-2000; 2000WO-US024008.
PF
XX
XX 03-SEP-1999; 99US-0152315P.
PR
XX
XX 03-SEP-1999; 99US-0152317P.
XX
XX (HUMA-) HUMAN GENOME SCT INC.
PA
XX

PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsu GA, Rosen CA,
PI Soppet DR, Young PG, Ebner R, Duan DR, Olsen HS, Lafleur DW,
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX MPI: 2001-203081/20.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
XX Disclosure: Page 18; 607pp; English.
XX
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX SQ Sequence 242 AA;
XX
XX Query Match 100.0%; Score 913; DB 4; Length 242;
XX Best Local Similarity 100.0%; Pred. No. 2.4e-94;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPDGTGALRASLVHVGSRPYTEPFPGQH 60
XX |||||
XX 59 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPDGTGALRASLVHVGSRPYTEPFPGQH 118
XX |||||
XX QY 61 SSGEAAQADAVRASACQMGDTHTGLALVYAKEQLFASASGARPGVPRVLYWVTDGSSDPV 120
XX |||||
XX DB 119 SSGEAAQADAVRASACQMGDTHTGLALVYAKEQLFASASGARPGVPRVLYWVTDGSSDPV 178
XX |||||
XX QY 121 GPPMOELKDLGVTFTVSTGRGNFLELSAASAPAKKHLFPVDVDDHLIIIVOLKRSIID 180
XX |||||
XX DB 179 GPPMOELKDLGVTFTVSTGRGNFLELSAASAPAKKHLFPVDVDDHLIIIVOLKRSIID 238
XX |||||
XX
XX RESULT 7
XX ID AAB88340 standard; protein; 285 AA.
XX AC AAB88340;
XX
XX 23-MAY-2001 (first entry)
XX
XX Human membrane or secretory protein clone P58C0053.
XX
XX Human; secretory protein; membrane protein; vaccine; gene therapy;
XX rheumatoid arthritis; diabetes.
KW

XX XX Homo sapiens.
XX OS
XX EP1067182-A2.
XX PN
XX 10-JAN-2001.
XX PD
XX 07-JUL-2000; 2000EP-00114090.
XX PF
XX 08-JUL-1999; 99JP-00194179.
XX PR 11-JAN-2000; 2000JP-00118775.
XX PR 02-MAY-2000; 2000JP-00183766.
XX
XX (HELI-1) HELIX RES INST.
XX PA
XX Ota T, Isegai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI MPI: 2001-093989/11.
XX DR N-PSDB; AAF93767.
XX
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development.
XX
XX Claim 1; SEQ ID NO 48; 609pp + Sequence Listing; English.
XX
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by AAB88317
CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
CC invention. The invention also includes methods for the production of
CC antibodies directed against the proteins, and cDNA sequences, which can
CC be used in vaccines. The polynucleotide sequences can be used in gene
CC therapy. The polynucleotide sequences and the proteins they encode may be
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate secretory protein/membrane protein expression. The
CC nucleic acids and complementary sequences may also be used as DNA probes
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
CC and quantitate the presence of similar nucleic acid sequences in samples.
CC They may also be used to study the expression and function of secretory
CC proteins/membrane polypeptides and their role in metabolism. The
CC polypeptides may be used as antigens in the production of antibodies
CC against them and in assays to identify modulators (agonists and
CC antagonists) of expression and activity. The antibodies and agonists
CC may also be used as therapeutic agents to down regulate expression and
CC activity. The antibodies may also be used as diagnostic agents for
CC detecting the presence of the polypeptides in samples (e.g. by enzyme
CC linked immunosorbent assay (ELISA). Examples of diseases which may be
CC treated include rheumatoid arthritis and diabetes
XX
XX SQ Sequence 285 AA;
XX
XX Query Match 100.0%; Score 913; DB 4; Length 285;
XX Best Local Similarity 100.0%; Pred. No. 3e-94;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPDGTGALRASLVHVGSRPYTEPFPGQH 60
XX |||||
XX DB 32 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPDGTGALRASLVHVGSRPYTEPFPGQH 91
XX |||||
XX QY 61 SSGEAAQADAVRASACQMGDTHTGLALVYAKEQLFASASGARPGVPRVLYWVTDGSSDPV 120
XX |||||
XX DB 92 SSGEAAQADAVRASACQMGDTHTGLALVYAKEQLFASASGARPGVPRVLYWVTDGSSDPV 151
XX |||||
XX QY 121 GPPMOELKDLGVTFTVSTGRGNFLELSAASAPAKKHLFPVDVDDHLIIIVOLKRSIID 180
XX |||||
XX DB 152 GPPMOELKDLGVTFTVSTGRGNFLELSAASAPAKKHLFPVDVDDHLIIIVOLKRSIID 211
XX |||||
XX
XX RESULT 8
XX ID ADY63045 standard; protein; 285 AA.
XX AC ADY63045;
XX

```
XX 02-JUN-2005 (first entry)
DT
XX Human clone PSEC0053 protein, SEQ ID 48.
DE
XX
XX Gene therapy.
KM
XX Homo sapiens.
OS
XX EPI514933-A1.
PN
XX 16-MAR-2005.
PD
XX 07-JUL-2000; 2004EP-00027228.
XX
XX 08-JUL-1999; 99JP-00194179.
XX 11-JAN-2000; 2000JP-00118775.
XX 02-MAY-2000; 2000JP-00183766.
XX 07-JUL-2000; 2000EP-00114090.
PR
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA
XX Ota T, Isegai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI WPI; 2005-203865/22.
DR N-PSDB; ADY63044.
XX
XX Novel isolated polynucleotide encoding human secretory proteins or
PT membrane proteins, useful for examination and diagnosis of abnormality of
PT human secretory proteins.
XX
XX Disclosure; SEQ ID NO 48; 1240pp; English.
PS
XX The present invention relates to novel human secretory proteins or
XX membrane proteins, and their coding sequences. The present sequence is
XX one such protein sequence. The coding sequences of the invention are
XX useful for examination and diagnosis of abnormality of the human
XX secretory proteins and in gene therapy methods. The coding sequences and
XX proteins are useful as candidates for medicines or as target molecules
XX for developing medicines. Antibodies against the proteins of the
XX invention are useful for treating diseases that are associated with the
XX proteins. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained from sequence information
XX supplied by the European Patent Office.
CC
XX
XX Sequence 285 AA;
SQ
XX
XX Query Match 100.0%; Score 913; DB 9; Length 285;
XX Best Local Similarity 100.0%; Pred. No. 3e-94;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGDLMFLDSSASVSHYEFSSRVREVGQLVAPLPGLGTGALRASLVHVGSRPYTEFPFGQH 60
OY
XX 32 RGDLMFLDSSASVSHYEFSSRVREVGQLVAPLPGLGTGALRASLVHVGSRPYTEFPFGQH 91
DB
XX 61 SSGEAAQDAVRASAGRMGDTHTGTLVYAKQLFAEASGARGPVKVLVWTTDGGSSDPV 120
OY
XX 92 SSGEAAQDAVRASAGRMGDTHTGTLVYAKQLFAEASGARGPVKVLVWTTDGGSSDPV 151
DB
XX 121 GPPMELKDGLGVTVIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
OY
XX 152 GPPMELKDGLGVTVIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 211
DB
XX
XX RESULT 9
XX AAE32502 standard; protein; 418 AA.
XX
XX AAE32502;
XX
XX 24-MAR-2003 (first entry)
DT
XX Human Willebrand Factor A domain related-protein (WARP).
DE
```

```
XX Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
KM extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
XX gene therapy; human.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..18
FT /label= signal_peptide
FT Protein 19..418
FT /note= "Human mature WARP protein"
FT Modified-site 148
FT /note= "O-glycosylation site"
FT Misc-difference 210..211
FT /note= "Encoded by CTCGCG"
FT Modified-site 264
FT /note= "N-glycosylation site"
FT Modified-site 359
FT /note= "N-glycosylation site"
FT Modified-site 361
FT /note= "O-glycosylation site"
FT Disulfide-bond 369..393
XX
XX WO200288184-A1.
XX
XX 07-NOV-2002.
XX
XX 02-MAY-2002; 2002WO-AU000542.
XX
XX 02-MAY-2001; 2001AU-00004701.
XX
XX (MURDOCH CHILDRENS RES INST.
XX
XX Bateman JF, Fitzgerald DJ;
XX WPI; 2003-111873/10.
XX N-PSDB; AAD50399.
XX
XX New isolated Willebrand Factor A-Related Protein polypeptide useful for
PT the manufacture of a medicament in the treatment of a disease condition
PT of the extracellular matrix, in particular arthritis.
XX
XX Claim 11; Page 76-78; 103pp; English.
PS
XX
XX The invention relates to Willebrand Factor A domain related-protein
XX (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
XX superfamily of extracellular matrix (ECM) proteins. WARP is used as a
XX molecular marker, used for detecting a loss of ECM integrity in an animal
XX subject, monitoring repair, regeneration or other disease processes in an
XX animal subject and detecting a disease condition or a propensity for the
XX development of a disease condition in an animal subject. The invention is
XX useful for the manufacture of a medicament in the treatment of a disease
XX condition of the ECM. The disease condition involves the cartilage, and
XX is preferably arthritis. The invention is also used in gene therapy. The
XX present sequence is human WARP protein
XX
XX Sequence 418 AA;
SQ
XX
XX Query Match 100.0%; Score 913; DB 6; Length 418;
XX Best Local Similarity 100.0%; Pred. No. 5.2e-94;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGDLMFLDSSASVSHYEFSSRVREVGQLVAPLPGLGTGALRASLVHVGSRPYTEFPFGQH 60
OY
XX 32 RGDLMFLDSSASVSHYEFSSRVREVGQLVAPLPGLGTGALRASLVHVGSRPYTEFPFGQH 91
DB
XX 61 SSGEAAQDAVRASAGRMGDTHTGTLVYAKQLFAEASGARGPVKVLVWTTDGGSSDPV 120
OY
XX 92 SSGEAAQDAVRASAGRMGDTHTGTLVYAKQLFAEASGARGPVKVLVWTTDGGSSDPV 151
DB
XX 121 GPPMELKDGLGVTVIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
OY
```


PR 19-JUN-2002; 2002US-0390006P.
 PR 19-JUN-2002; 2002US-0390209P.
 PR 21-JUN-2002; 2002US-0390763P.
 PR 17-JUL-2002; 2002US-0396706P.
 PR 06-AUG-2002; 2002US-0401628P.
 PR 09-AUG-2002; 2002US-0402155P.
 PR 09-AUG-2002; 2002US-0402256P.
 PR 09-AUG-2002; 2002US-0402389P.
 PR 12-AUG-2002; 2002US-0402816P.
 PR 12-AUG-2002; 2002US-0402821P.
 PR 12-AUG-2002; 2002US-0402832P.
 PR 13-AUG-2002; 2002US-0403448P.
 PR 13-AUG-2002; 2002US-0403459P.
 PR 13-AUG-2002; 2002US-0403531P.
 PR 13-AUG-2002; 2002US-0403532P.
 PR 13-AUG-2002; 2002US-0403563P.
 PR 15-AUG-2002; 2002US-0406317P.
 PR 15-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406355P.
 PR 27-AUG-2002; 2002US-0406240P.
 PR 12-SEP-2002; 2002US-0410084P.
 PR 20-SEP-2002; 2002US-0412523P.
 PR 23-SEP-2002; 2002US-041731P.
 PR 30-SEP-2002; 2002US-041731P.
 PR 30-SEP-2002; 2002US-0418039P.
 PR 30-SEP-2002; 2002US-041840P.
 PR 30-SEP-2002; 2002US-041954P.
 PR 09-OCT-2002; 2002US-0417186P.
 PR 09-OCT-2002; 2002US-0417406P.
 PR 23-OCT-2002; 2002US-0420639P.
 PR 28-OCT-2002; 2002US-0421156P.
 PR 31-OCT-2002; 2002US-0422690P.
 PR 01-NOV-2002; 2002US-0423130P.
 PR 05-NOV-2002; 2002US-00423798P.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 12-NOV-2002; 2002US-0425453P.

(CURA-) CURAGEN CORP.

XX PA
 PI Alabrook JP, Alvarez E, Anderson DM, Boldog FL, Casman SJ;
 PI Catterton E, Chappoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
 PI Eitenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
 PI Gusev VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, Macdougall JR;
 PI MacLachlan T, Malysankar UM, Mezick AJ, Miller I, Mishra VS;
 PI Padigaru M, Patrajan M, Pena CE, Peyman JA, Raha D, Rastrelli L;
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SC, Shinkets RA;
 PI Smitson G, Spytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M;
 PI Zhong H;
 XX
 DR WPI; 2004-081935/08.
 DR N-PSDB; ADH71105.
 XX
 PT New NOVA polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVA-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 XX
 PS Example 1; SEQ ID NO 2; 1880pp; English.

XX CC The invention relates to a novel isolated polypeptide (NOVA). A
 CC polypeptide of the invention has cytostatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antihypertensive activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOVA polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing,
 CC treating or preventing NOVA-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are
 CC further used as hybridisation probes, in chromosome mapping, tissue

CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC represents a NOVA polypeptide of the invention.
 XX
 SQ Sequence 445 AA;

Query Match 100.0%; Score 913; DB 8; Length 445;
 Best Local Similarity 100.0%; Pred. No. 5.7e-94;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHFEFSRREFGQLVPLPLGICALASLVHVSRRPYTEFPFGQH 60
 DB 32 RGDLMFLDSSASVSHFEFSRREFGQLVPLPLGICALASLVHVSRRPYTEFPFGQH 91
 QY 61 SSGEAAQDAVAASQRMGDTHTGALVYAKQLFAEASGAPGVKVLVWTTDGGSSDPV 120
 DB 92 SSGEAAQDAVAASQRMGDTHTGALVYAKQLFAEASGAPGVKVLVWTTDGGSSDPV 151
 QY 121 GPMQELKDLGVTVFIVSTGNGFLESAASAPAEKHLHFVDVDDLIITVQELRGSLTD 180
 DB 152 GPMQELKDLGVTVFIVSTGNGFLESAASAPAEKHLHFVDVDDLIITVQELRGSLTD 211

RESULT 12

AAE03654
 ID AAE03654 standard; protein; 185 AA.

AAE03654;

06-AUG-2001 (first entry)

Human extracellular matrix and cell adhesion molecule-18 (XMAD-18).

KW Human; extracellular matrix and cell adhesion molecule; XMAD;
 KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
 KW sickle cell anaemia; thalassemia; autoimmune disorder; adenocarcinoma;
 KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
 KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
 KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
 KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
 KW infection; cell proliferative disorder; actinic keratosis; myeloma;
 KW arteriosclerosis; nootropic; anticonvulsant; antithyroid; nephrotoxic;
 KW neuroprotective; dermatological.

Homosapiens.

OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT 1..171
 FT /note= "Collagen glycoprotein precursor"
 FT Region
 FT 1..170
 FT /note= "Von Willebrand factor domain score"
 FT Region
 FT 2..15
 FT /note= "Collagen glycoprotein precursor"
 FT Region
 FT 37..51
 FT /note= "Collagen glycoprotein precursor"
 FT Region
 FT 103..111
 FT /note= "Collagen glycoprotein precursor"

W0200142285-A2.

14-JUN-2001.

05-DEC-2000; 2000WO-US032990.

10-DEC-1999; 99US-0172852P.

16-DEC-1999; 99US-0172354P.

(INCY-) INCYTE GENOMICS INC.

Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;

Baughn MR, Lu DM, Shah P, Au-Young J;

DR MPI, 2001-381632/40.
 XX N-PSDB; AAD08062.
 XX New human extracellular matrix and cell adhesion molecules and
 PT polynucleotide sequences encoding them, useful for diagnosis, prevention,
 PT treatment of genetic, autoimmune and cell proliferative disorders.
 XX
 PS Claim 1; Page 111; 135pp; English.
 XX
 CC The present sequence is a human extracellular matrix and cell adhesion
 CC molecule (XMAD). The XMAD is used for screening a compound for
 CC effectiveness as an agonist or antagonist of XMAD. The identified agonist
 CC or antagonist are used for treating a disease or condition associated
 CC with decreased or increased expression of functional XMAD. The
 CC polynucleotides encoding XMAD are useful in somatic or germline gene
 CC therapy to correct a genetic deficiency, to express a conditionally
 CC lethal gene product and to express a protein which affords protection
 CC against intracellular parasites and also for diagnosis of disorders
 CC associated with expression of XMAD. They are also used for generating
 CC hybridisation probes useful in mapping the naturally occurring genomic
 CC sequences and to create knock in humanised animals (pigs) or transgenic
 CC animals (mice or rats) to model human diseases. Oligonucleotide or longer
 CC fragments derived from the polynucleotide sequences may be used as
 CC elements on a microarray. Antibodies which specifically bind XMAD may be
 CC used for the diagnosis of disorders associated with the expression of
 CC XMAD, or in assays to monitor patients being treated with XMAD. Diseases
 CC diagnosed, prevented or treated include genetic disorders such as
 CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 CC disease, myotonic dystrophy, sickle cell anaemia, thalassemia,
 CC autoimmune/inflammatory disorders such as acquired immune deficiency
 CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
 CC atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
 CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 CC bacterial, fungal, parasitic, protozoal and helminthic infections and
 CC cell proliferative disorders such as actinic keratosis, arteriosclerosis
 CC and cancer including breast, bladder, bone marrow, brain and uterus
 CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma
 CC
 XX Sequence 185 AA;
 SQ
 Query Match 97.7%; Score 892; DB 4; Length 185;
 Best Local Similarity 100.0%; Pred. No. 3.9e-92;
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 5 MFLDSSASVSHSEFRVREPVQGLVAPLPLGIGARASLVNHGSRPYTEPPFGQHSSE 64
 Db 1 MFLDSSASVSHSEFRVREPVQGLVAPLPLGIGARASLVNHGSRPYTEPPFGQHSSE 60
 Oy 65 AADAVRASAGNRGDTHTGTALVYAKEQLFAEASGARPGVPKVLWVWTDGSSDPVGP 124
 Db 61 AADAVRASAGNRGDTHTGTALVYAKEQLFAEASGARPGVPKVLWVWTDGSSDPVGP 120
 Oy 125 QELKDLGVTVFIYVSTGRGNFLELSAASAPAEKHLHFVVDVDDLIHIVQELRGSLD 180
 Db 121 QELKDLGVTVFIYVSTGRGNFLELSAASAPAEKHLHFVVDVDDLIHIVQELRGSLD 176

KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US0008621.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinketsu RA, Leach M;
 XX
 DR MPI; 2000-602362/57.
 XX N-PSDB; AAC76790.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 11; Page 3880-3881; 5507pp; English.
 XX
 CC AAC7446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; neuroprotective; osteoprotic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancer, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 CC
 XX Sequence 299 AA;
 SQ
 Query Match 96.7%; Score 883; DB 3; Length 299;
 Best Local Similarity 98.3%; Pred. No. 8e-91;
 Matches 175; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 3 DLMFLDSSASVSHSEFRVREPVQGLVAPLPLGIGARASLVNHGSRPYTEPPFGQHS 62
 Db 37 DLMFLDSSASVSHSEFRVREPVQGLVAPLPLAPALASLVNHGSRPYTEPPFGQHS 96
 Oy 63 GEAADAVRASAGNRGDTHTGTALVYAKEQLFAEASGARPGVPKVLWVWTDGSSDPVGP 122
 Db 97 GEAADAVRASAGNRGDTHTGTALVYAKEQLFAEASGARPGVPKVLWVWTDGSSDPVGP 156
 Oy 123 PEOELKDLGVTVFIYVSTGRGNFLELSAASAPAEKHLHFVVDVDDLIHIVQELRGSLD 180

Db 157 PMQELKDGTVTVIVSTGRGNFLELSAASAPAEKHLHFVDVDDLIHTVQELRGSILD 214

```
RESULT 14
AAE32503
ID AAE32503 standard; protein; 180 AA.
XX
XX AAE32503;
XX
XX 24-MAR-2003 (first entry)
XX
XX Mouse von Willebrand Factor A (VA) domain.
XX
XX Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
XX extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
XX gene therapy; mouse.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX Misc-difference 1..2
XX /note= "Encoded by GGG"
XX
XX WO200288184-A1.
XX
XX 07-NOV-2002.
XX
XX 02-MAY-2002; 2002WO-AU000542.
XX
XX 02-MAY-2001; 2001AU-00004701.
XX
XX (MURD-) MURDOCH CHILDRENS RES INST.
XX
XX Bateman JF, Fitzgerald DJ;
XX
XX WPI; 2003-111873/10.
XX
XX N-PSDB; AAD50400.
XX
XX New isolated Willebrand Factor A-Related Protein polypeptide useful for
XX the manufacture of a medicament in the treatment of a disease condition
XX of the extracellular matrix, in particular arthritis.
XX
XX Claim 9; Page 78-79; 103pp; English.
XX
XX The invention relates to Willebrand Factor A domain related-protein
XX (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
XX superfamily of extracellular matrix (ECM) proteins. WARP is used as a
XX molecular marker, used for detecting a loss of ECM integrity in an animal
XX subject, monitoring repair, regeneration or other disease processes in an
XX animal subject and detecting a disease condition or a propensity for the
XX development of a disease condition in an animal subject. The invention is
XX useful for the manufacture of a medicament in the treatment of a disease
XX condition of the ECM. The disease condition involves the cartilage, and
XX is preferably arthritis. The invention is also used in gene therapy. The
XX present sequence is mouse VA domain
XX
XX Sequence 180 AA;
```

Query Match 84.1%; Score 768; DB 6; Length 180;
Best Local Similarity 83.3%; Pred. No. 4e-76; Mismatches 16; Indels 0; Gaps 0;

```
Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;
Oy 1 RGDLMFLDSSASVSHYFERSVRREFVGLVAPLPLGTGALRASLVHVGSRPYTFPPFGQH 60
Db 1 QGDLLFLDSSASVSHYFERSVRREFVGLVATMSFGPLRLSLVHVGSPHTFTFDY 60
Oy 61 SSGEAAQDAVRASAO RMGDTHTGLALVYAKQLFAEASGARPVPKVLVWTTGGSSDPV 120
Db 61 SSGQAIRDAIRVAPQRMGDTHTGLALVYAKQLFAEAGARPVPKVLVWTTGGSSDPV 120
Oy 121 GPMQELKDLGVTVTVIVSTGRGNFLELSAASAPAEKHLHFVDVDDLIHTVQELRGSILD 180
Db 121 GPMQELKDLGVTVTVIVSTGRGNLLELLAASAPAEKHLHFVDVDDLIHTVQELRGSILD 180
```

```
RESULT 15
AAE32501
ID AAE32501 standard; protein; 415 AA.
XX
XX AAE32501;
XX
XX 24-MAR-2003 (first entry)
XX
XX Mouse Willebrand Factor A domain related-protein (WARP).
XX
XX Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
XX extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
XX gene therapy; mouse.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX Peptide 1..18
XX /label= Signal_peptide
XX Protein 19..415
XX /note= "Mouse mature WARP protein"
XX
XX Modified-site 148
XX /note= "O-glycosylation site"
XX
XX Modified-site 264
XX /note= "N-glycosylation site"
XX
XX Modified-site 359
XX /note= "N-glycosylation site"
XX
XX Modified-site 361
XX /note= "O-glycosylation site"
XX
XX Disulfide-bond 369..393
XX Modified-site 400
XX /note= "O-glycosylation site"
XX
XX WO200288184-A1.
XX
XX 07-NOV-2002.
XX
XX 02-MAY-2002; 2002WO-AU000542.
XX
XX 02-MAY-2001; 2001AU-00004701.
XX
XX (MURD-) MURDOCH CHILDRENS RES INST.
XX
XX Bateman JF, Fitzgerald DJ;
XX
XX WPI; 2003-111873/10.
XX
XX N-PSDB; AAD50022; AAD50398.
XX
XX New isolated Willebrand Factor A-Related Protein polypeptide useful for
XX the manufacture of a medicament in the treatment of a disease condition
XX of the extracellular matrix, in particular arthritis.
XX
XX Claim 13; Page 74-75; 103pp; English.
XX
XX The invention relates to Willebrand Factor A domain related-protein
XX (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
XX superfamily of extracellular matrix (ECM) proteins. WARP is used as a
XX molecular marker, used for detecting a loss of ECM integrity in an animal
XX subject, monitoring repair, regeneration or other disease processes in an
XX animal subject and detecting a disease condition or a propensity for the
XX development of a disease condition in an animal subject. The invention is
XX useful for the manufacture of a medicament in the treatment of a disease
XX condition of the ECM. The disease condition involves the cartilage, and
XX is preferably arthritis. The invention is also used in gene therapy. The
XX present sequence is mouse WARP protein
XX
XX Sequence 415 AA;
```

Query Match 84.1%; Score 768; DB 6; Length 415;
Best Local Similarity 83.3%; Pred. No. 1.3e-77; Mismatches 16; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 13:10:29 ; Search time 36.7224 Seconds
(without alignments)
2048.049 Million cell updates/sec

Title: US-10-699-035A-2

Perfect score: 913
Sequence: 1 RQDLMFLDSSASVSHYFES.....FVDDVDDHIIIVOLRGSIID 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_Main:*
1: /cgn2_6/prodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubppaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubppaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubppaa/US10_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubppaa/US10_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubppaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	913	100.0	180	4	US-10-699-035A-2
2	913	100.0	215	3	US-09-789-561-85
3	913	100.0	215	3	US-09-833-245-2096
4	913	100.0	215	5	US-10-883-936-85
5	913	100.0	226	3	US-09-789-561-165
6	913	100.0	226	5	US-10-883-936-165
7	913	100.0	242	3	US-09-789-561-159
8	913	100.0	242	5	US-10-883-936-159
9	913	100.0	418	4	US-10-699-035A-6
10	913	100.0	418	4	US-10-699-035A-20
11	892	97.7	185	4	US-10-149-819-18
12	892	97.7	185	4	US-10-699-035A-8
13	84.1	84.1	181	4	US-10-699-035A-31
14	768	84.1	415	4	US-10-699-035A-4
15	768	84.1	415	4	US-10-699-035A-21
16	272.5	29.8	176	3	US-09-976-782-38
17	265.5	29.8	182	4	US-10-699-035A-22
18	263.5	28.9	184	4	US-10-699-035A-27
19	263.5	28.9	3063	3	US-09-918-715-257
20	263.5	28.9	3063	4	US-10-177-293-61
21	263.5	28.9	3063	4	US-10-177-293-61
22	263.5	28.9	3063	4	US-10-301-822-26
23	263.5	28.9	3063	4	US-10-474-794-257
24	263.5	28.9	3063	5	US-10-979-159-257
25	263.5	28.9	3063	5	US-10-631-467-918
26	263.5	28.9	3118	5	US-10-220-335-287
27	259	28.4	580	5	US-10-723-860-2769

28	256.5	28.1	755	3	US-09-919-497-57	Sequence 57, Appl
29	256.5	28.1	1280	4	US-10-115-479-10	Sequence 10, Appl
30	254.5	27.9	1207	4	US-10-408-765A-1591	Sequence 1591, Ap
31	254.5	27.9	1297	4	US-10-187-975-102	Sequence 102, Appl
32	251.5	27.5	776	4	US-10-000-512-8	Sequence 8, Appl
33	251.5	27.5	776	4	US-10-074-556-8	Sequence 176, Appl
34	251.5	27.5	782	4	US-10-428-275-176	Sequence 1292, Ap
35	251.5	27.5	794	3	US-09-833-245-1592	Sequence 16, Appl
36	251.5	27.5	896	4	US-10-004-378A-16	Sequence 158, App
37	251.5	27.5	896	4	US-10-428-275-158	Sequence 156, App
38	251.5	27.5	915	3	US-10-428-275-156	Sequence 34, Appl
39	251.5	27.5	915	3	US-09-909-320-34	Sequence 34, Appl
40	251.5	27.5	915	3	US-09-903-291A-34	Sequence 34, Appl
41	251.5	27.5	915	3	US-09-902-853-34	Sequence 34, Appl
42	251.5	27.5	915	3	US-09-907-824-34	Sequence 34, Appl
43	251.5	27.5	915	3	US-09-907-841-34	Sequence 34, Appl
44	251.5	27.5	915	3	US-09-904-011-34	Sequence 34, Appl
45	251.5	27.5	915	3	US-09-904-011-34	Sequence 34, Appl

ALIGNMENTS

```

RESULT 1
US-10-699-035A-2
; Sequence 2, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-2

Query Match      100.0%; Score 913; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.5e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RQDLMFLDSSASVSHYFESRVREFYQGLVADPLQGTALRASLVHVGSRPYTEPPFGQH 60
        |||||||
Db       1 RQDLMFLDSSASVSHYFESRVREFYQGLVADPLQGTALRASLVHVGSRPYTEPPFGQH 60

QY      61 SSGEAAQDAVRASQRMGDTHTGLALVYAKEOLFPAASGARGPVKVLVWTDGSSDPV 120
        |||||||
Db       61 SSGEAAQDAVRASQRMGDTHTGLALVYAKEOLFPAASGARGPVKVLVWTDGSSDPV 120

QY      121 GPMQGLKQLGTVTVFVSTGRNPFLELSAASAPAEKHLHFPDVDDHLIIIVOLRGSIID 180
        |||||||
Db       121 GPMQGLKQLGTVTVFVSTGRNPFLELSAASAPAEKHLHFPDVDDHLIIIVOLRGSIID 180

RESULT 2
US-09-789-561-85
; Sequence 85, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P20433P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22

```

```

; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-85

Query Match          100.0%; Score 913; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.9e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 60
Db 32 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 91
Qy 61 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 120
Db 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
Qy 121 GPPMOELKDGLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
Db 152 GPPMOELKDGLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 211

RESULT 3
US-09-833-245-2096
; Sequence 2096, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PFS46PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2096
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2096

Query Match          100.0%; Score 913; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.9e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 60
Db 32 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 91
Qy 61 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 120
Db 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
Qy 121 GPPMOELKDGLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
Db 152 GPPMOELKDGLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 211
```

```

Db 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
Qy 121 GPPMOELKDGLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
Db 152 GPPMOELKDGLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 211

RESULT 4
US-10-883-936-85
; Sequence 85, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; PRIOR FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-85

Query Match          100.0%; Score 913; DB 5; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.9e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 60
Db 32 RGDLMFLDSSASVSHYEFRRVREVGQVAPLPGLTGALRASLVHVGSRPYTEPPFGQH 91
Qy 61 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 120
Db 92 SSGEAAQDAVRAAQRMGDTHTGLALVYAKQLFAEASGARPGVPKVLVWTTDGGSSDPV 151
Qy 121 GPPMOELKDGLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 180
Db 152 GPPMOELKDGLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDDDLHIIVQELRGSLTD 211

RESULT 5
US-09-789-561-165
; Sequence 165, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 165
```

LENGTH: 226
TYPE: PRT
ORGANISM: Homo sapiens
US-09-789-561-159

Query Match 100.0%; Score 913; DB 3; Length 226;
Best Local Similarity 100.0%; Pred. No. 2e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYFSRREFFVGQLVAPLPCTGALRASLVHVGSRPTEPPFGQH 60
DB 45 RGDLMFLDSSASVSHYFSRREFFVGQLVAPLPCTGALRASLVHVGSRPTEPPFGQH 104
QY 61 SSGEAAQDAVRASAGRMGDTHTGLALVYAKEQLFAEASGARPGVPCVLVWVTGGSSDPV 120
DB 105 SSGEAAQDAVRASAGRMGDTHTGLALVYAKEQLFAEASGARPGVPCVLVWVTGGSSDPV 164
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 180
DB 165 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 224

RESULT 6
US-10-883-936-165
Sequence 165, Application US/10883936
Publication No. US20050019866A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: PZ043P1
CURRENT APPLICATION NUMBER: US/10/883,936
CURRENT FILING DATE: 2004-07-06
PRIOR APPLICATION NUMBER: US/09/789,561
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 165
LENGTH: 226
TYPE: PRT
ORGANISM: Homo sapiens
US-10-883-936-165

Query Match 100.0%; Score 913; DB 5; Length 226;
Best Local Similarity 100.0%; Pred. No. 2e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYFSRREFFVGQLVAPLPCTGALRASLVHVGSRPTEPPFGQH 60
DB 45 RGDLMFLDSSASVSHYFSRREFFVGQLVAPLPCTGALRASLVHVGSRPTEPPFGQH 104
QY 61 SSGEAAQDAVRASAGRMGDTHTGLALVYAKEQLFAEASGARPGVPCVLVWVTGGSSDPV 120
DB 105 SSGEAAQDAVRASAGRMGDTHTGLALVYAKEQLFAEASGARPGVPCVLVWVTGGSSDPV 164
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 180
DB 165 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 224

RESULT 7
US-09-789-561-159
Sequence 159, Application US/09789561
Patent No. US20020064818A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: PZ043P1

CURRENT APPLICATION NUMBER: US/09/789,561
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 159
LENGTH: 242
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-159

Query Match 100.0%; Score 913; DB 3; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.2e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYFSRREFFVGQLVAPLPCTGALRASLVHVGSRPTEPPFGQH 60
DB 59 RGDLMFLDSSASVSHYFSRREFFVGQLVAPLPCTGALRASLVHVGSRPTEPPFGQH 118
QY 61 SSGEAAQDAVRASAGRMGDTHTGLALVYAKEQLFAEASGARPGVPCVLVWVTGGSSDPV 120
DB 119 SSGEAAQDAVRASAGRMGDTHTGLALVYAKEQLFAEASGARPGVPCVLVWVTGGSSDPV 178
QY 121 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 180
DB 179 GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSILD 238

RESULT 8
US-10-883-936-159

Sequence 159, Application US/10883936
Publication No. US20050019866A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 52 Human secreted proteins
FILE REFERENCE: PZ043P1
CURRENT APPLICATION NUMBER: US/10/883,936
CURRENT FILING DATE: 2004-07-06
PRIOR APPLICATION NUMBER: US/09/789,561
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: PCT/US00/24008
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/152,317
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/152,315
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 159
LENGTH: 242
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE

LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-159

Query Match 100.0%; Score 913; DB 5; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.2e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 60
DB 59 RGDLMFLDSSASVSHYEFSSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 118
QY 61 SSGEAAQDAVARASQRMGDTHTGTLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 120
DB 119 SSGEAAQDAVARASQRMGDTHTGTLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 178
QY 121 GPPMOELKDGLGVTVPFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
DB 179 GPPMOELKDGLGVTVPFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 238

RESULT 9
US-10-699-035A-6
Sequence 6, Application US/10699035A
Publication No. US20040214349A1
GENERAL INFORMATION:
APPLICANT: Bateman, John
APPLICANT: Fitzgerald, David
TITLE OF INVENTION: A Molecular Marker
FILE REFERENCE: A36056 PCT USA A 071838.0142
CURRENT APPLICATION NUMBER: US/10/699,035A
CURRENT FILING DATE: 2003-10-31
PRIOR APPLICATION NUMBER: PCT/AU02/00542
PRIOR FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: AU PR4701/01
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 418
TYPE: PRT
ORGANISM: Homo sapiens
US-10-699-035A-6

Query Match 100.0%; Score 913; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 4.7e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFSSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
QY 61 SSGEAAQDAVARASQRMGDTHTGTLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVARASQRMGDTHTGTLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 151
QY 121 GPPMOELKDGLGVTVPFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
DB 152 GPPMOELKDGLGVTVPFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211

RESULT 10
US-10-699-035A-20
Sequence 20, Application US/10699035A
Publication No. US20040214349A1
GENERAL INFORMATION:
APPLICANT: Bateman, John
APPLICANT: Fitzgerald, David
TITLE OF INVENTION: A Molecular Marker

FILE REFERENCE: A36056 PCT USA A 071838.0142
CURRENT APPLICATION NUMBER: US/10/699,035A
CURRENT FILING DATE: 2003-10-31
PRIOR APPLICATION NUMBER: PCT/AU02/00542
PRIOR FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: AU PR4701/01
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 418
TYPE: PRT
ORGANISM: Homo sapiens
US-10-699-035A-20

Query Match 100.0%; Score 913; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 4.7e-89;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 60
DB 32 RGDLMFLDSSASVSHYEFSSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
QY 61 SSGEAAQDAVARASQRMGDTHTGTLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 120
DB 92 SSGEAAQDAVARASQRMGDTHTGTLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 151
QY 121 GPPMOELKDGLGVTVPFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
DB 152 GPPMOELKDGLGVTVPFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211

RESULT 11
US-10-149-819-18
Sequence 18, Application US/10149819
Publication No. US20030044913A1
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: YUE, Henry
APPLICANT: AZIMZAI, Yalda
APPLICANT: TANG, Y. Tom
APPLICANT: PATTERSON, Chandra
APPLICANT: BAUGHN, Mariah R.
APPLICANT: LU, Dyung Aina M.
APPLICANT: SHAH, Pooji
APPLICANT: IAL, Preci
APPLICANT: AU-YOUNG, Janice
APPLICANT: BURFORD, Neil
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
FILE REFERENCE: PF-0760 PCT
CURRENT APPLICATION NUMBER: US/10/149,819
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
PRIOR FILING DATE: 1999-12-10; 1999-12-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PERL Program
SEQ ID NO 18
LENGTH: 185
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030044913A1 3683905CD1
US-10-149-819-18

Query Match 97.7%; Score 892; DB 4; Length 185;
Best Local Similarity 100.0%; Pred. No. 2.7e-87;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 MFLDSSASVSHYEFSSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQSSGE 64
DB 1 MFLDSSASVSHYEFSSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEPPFGQSSGE 60

QY 65 AADAVRASAOQMGDTHTGLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPVPGPM 124
 DB 61 AADAVRASAOQMGDTHTGLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPVPGPM 120
 QY 125 QELKDLGVTIVSTGRGNFLELSAASAPAEKHLHFVVDVDDLHIIVDELKSGSIID 180
 DB 121 QELKDLGVTIVSTGRGNFLELSAASAPAEKHLHFVVDVDDLHIIVDELKSGSIID 176

RESULT 12

US-10-699-035a-8
 ; Sequence 8, Application US/10699035A
 ; Publication No. US20040214349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bateman, John
 ; APPLICANT: Fitzgerald, David
 ; TITLE OF INVENTION: A Molecular Marker
 ; FILE REFERENCE: A36056 PCT USA A 071838.0142
 ; CURRENT APPLICATION NUMBER: US/10/699,035A
 ; CURRENT FILING DATE: 2003-10-31
 ; PRIOR APPLICATION NUMBER: PCT/AU02/00542
 ; PRIOR FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: AU PR4701/01
 ; PRIOR FILING DATE: 2001-05-02
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 180
 ; TYPE: PRP
 ; ORGANISM: Mus musculus
 US-10-699-035a-8

Query Match 84.1%; Score 768; DB 4; Length 180;
 Best Local Similarity 83.3%; Pred. No. 5.5e-74;
 Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVSRRPYTEPPFGQH 60
 DB 1 QGDLMFLDSSASVSHYEFSSRVREFVQGLVATMSFGPALRASLVHVSQPHTEFTFDQY 60
 QY 61 SSGEAAODAVRASAOQMGDTHTGLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 120
 DB 61 SSGQAIRDAIRVAPQRMGDTHTGLALVYAKEQLFAEAGARGVPKVLVWTTDGGSSDPV 120
 QY 121 GPPMOELKDLGVTIVSTGRGNFLELSAASAPAEKHLHFVVDVDDLHIIVDELKSGSIID 180
 DB 121 GPPMOELKDLGVTIVSTGRGNFLELSAASAPAEKHLHFVVDVDDLHIIVDELKSGSIID 180

RESULT 13

US-10-699-035a-31
 ; Sequence 31, Application US/10699035A
 ; Publication No. US20040214349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bateman, John
 ; APPLICANT: Fitzgerald, David
 ; TITLE OF INVENTION: A Molecular Marker
 ; FILE REFERENCE: A36056 PCT USA A 071838.0142
 ; CURRENT APPLICATION NUMBER: US/10/699,035A
 ; CURRENT FILING DATE: 2003-10-31
 ; PRIOR APPLICATION NUMBER: PCT/AU02/00542
 ; PRIOR FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: AU PR4701/01
 ; PRIOR FILING DATE: 2001-05-02
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 31
 ; LENGTH: 181
 ; TYPE: PRP
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: VA domain from WARP
 US-10-699-035a-31

Query Match 84.1%; Score 768; DB 4; Length 181;
 Best Local Similarity 83.3%; Pred. No. 5.5e-74;
 Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVSRRPYTEPPFGQH 60
 DB 1 QGDLMFLDSSASVSHYEFSSRVREFVQGLVATMSFGPALRASLVHVSQPHTEFTFDQY 60
 QY 61 SSGEAAODAVRASAOQMGDTHTGLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 120
 DB 61 SSGQAIRDAIRVAPQRMGDTHTGLALVYAKEQLFAEAGARGVPKVLVWTTDGGSSDPV 120
 QY 121 GPPMOELKDLGVTIVSTGRGNFLELSAASAPAEKHLHFVVDVDDLHIIVDELKSGSIID 180
 DB 121 GPPMOELKDLGVTIVSTGRGNFLELSAASAPAEKHLHFVVDVDDLHIIVDELKSGSIID 180

RESULT 14

US-10-699-035a-4
 ; Sequence 4, Application US/10699035A
 ; Publication No. US20040214349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bateman, John
 ; APPLICANT: Fitzgerald, David
 ; TITLE OF INVENTION: A Molecular Marker
 ; FILE REFERENCE: A36056 PCT USA A 071838.0142
 ; CURRENT APPLICATION NUMBER: US/10/699,035A
 ; CURRENT FILING DATE: 2003-10-31
 ; PRIOR APPLICATION NUMBER: PCT/AU02/00542
 ; PRIOR FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: AU PR4701/01
 ; PRIOR FILING DATE: 2001-05-02
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 415
 ; TYPE: PRP
 ; ORGANISM: Mus musculus
 US-10-699-035a-4

Query Match 84.1%; Score 768; DB 4; Length 415;
 Best Local Similarity 83.3%; Pred. No. 1.7e-73;
 Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

QY 1 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVSRRPYTEPPFGQH 60
 DB 32 QGDLMFLDSSASVSHYEFSSRVREFVQGLVATMSFGPALRASLVHVSQPHTEFTFDQY 91
 QY 61 SSGEAAODAVRASAOQMGDTHTGLALVYAKEQLFAEASGARGVPKVLVWTTDGGSSDPV 120
 DB 92 SSGQAIRDAIRVAPQRMGDTHTGLALVYAKEQLFAEAGARGVPKVLVWTTDGGSSDPV 151
 QY 121 GPPMOELKDLGVTIVSTGRGNFLELSAASAPAEKHLHFVVDVDDLHIIVDELKSGSIID 180
 DB 152 GPPMOELKDLGVTIVSTGRGNFLELSAASAPAEKHLHFVVDVDDLHIIVDELKSGSIID 211

RESULT 15

US-10-699-035a-21
 ; Sequence 21, Application US/10699035A
 ; Publication No. US20040214349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bateman, John
 ; APPLICANT: Fitzgerald, David
 ; TITLE OF INVENTION: A Molecular Marker
 ; FILE REFERENCE: A36056 PCT USA A 071838.0142
 ; CURRENT APPLICATION NUMBER: US/10/699,035A
 ; CURRENT FILING DATE: 2003-10-31
 ; PRIOR APPLICATION NUMBER: PCT/AU02/00542
 ; PRIOR FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: AU PR4701/01
 ; PRIOR FILING DATE: 2001-05-02

; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 21
 ; LENGTH: 415
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-699-035A-21

Query Match 84.1%; Score 768; DB 4; Length 415;
 Best Local Similarity 83.3%; Pred. No. 1.7e-73;
 Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

Qy	1	RGDLMTLLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTERFPQGH	60
Db	32	QGDLFLFLDSSASVSHYEFSSRVREFVQGLVATMSPGALRASLVHVGSPHTTEFTDQY	91
Qy	61	SSGEAADAVRASAORMDTHTGTLAVYAKEQLFAEASGARPGVPKVLVWVTDDGSSDPV	120
Db	92	SSGQAIRDAIRVAPQRMGDTNTGLALAYAKEQLFAEBAGARPGVPKVLVWVTDDGSSDPV	151
Qy	121	GPPMOELKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSITD	180
Db	152	GPPMOELKDLGVTVFIVSTGRGNLLELSAASAPAEKHLHFVDVDDLPIARELRGSITD	211

Search completed: February 13, 2006, 13:12:39
 Job time : 37.7224 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: February 13, 2006, 07:47:59 ; Search time 27.6923 Seconds
(without alignments)
537.392 Million cell updates/sec

Title: US-10-699-035A-2

Perfect score: 913
Sequence: 1 RGDMLFLDSASVSHYERS.....FVDYDHLHIVQELRGLD 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/ECTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	384	42.1	77	US-09-513-999C-6277	Sequence 6277, Ap
2	259	28.4	584	US-09-949-016-10340	Sequence 10340, A
3	259	28.4	584	US-09-949-016-10341	Sequence 10341, A
4	256.5	28.1	755	US-09-919-497-57	Sequence 57, Appl
5	251.5	27.5	638	US-08-897-443-1	Sequence 1, Appl
6	251.5	27.5	915	US-09-907-794A-34	Sequence 34, Appl
7	251.5	27.5	915	US-09-905-125A-34	Sequence 34, Appl
8	251.5	27.5	915	US-09-902-775A-34	Sequence 34, Appl
9	251.5	27.5	915	US-09-906-700-34	Sequence 34, Appl
10	251.5	27.5	915	US-09-903-603A-34	Sequence 34, Appl
11	251.5	27.5	915	US-09-904-920A-34	Sequence 34, Appl
12	251.5	27.5	915	US-09-909-064-34	Sequence 34, Appl
13	251.5	27.5	915	US-09-905-381A-34	Sequence 34, Appl
14	251.5	27.5	915	US-09-906-618-34	Sequence 34, Appl
15	251.5	27.5	915	US-09-906-646-34	Sequence 34, Appl
16	251.5	27.5	915	US-09-904-462-34	Sequence 34, Appl
17	251.5	27.5	915	US-09-902-736A-34	Sequence 34, Appl
18	251.5	27.5	915	US-09-906-722A-34	Sequence 34, Appl
19	251.5	27.5	915	US-09-949-016-6215	Sequence 6215, Ap
20	251.5	27.5	915	US-09-949-016-11519	Sequence 11519, A
21	251.5	27.5	915	US-09-949-016-11519	Sequence 11519, A
22	242.5	26.6	956	US-08-897-443-3	Sequence 3, Appl
23	233.5	25.6	496	US-08-462-128-37	Sequence 37, Appl
24	233.5	25.6	496	US-08-463-180-37	Sequence 37, Appl
25	233.5	25.6	496	US-08-001-078A-1	Sequence 1, Appl
26	233.5	25.6	496	US-08-897-443-4	Sequence 4, Appl
27	233.5	25.6	496	US-08-463-218-1	Sequence 1, Appl

28	233.5	25.6	496	2	US-09-949-016-11306	Sequence 11306, A
29	233.5	25.6	496	4	PCT-US94-00253-1	Sequence 1, Appl
30	231.5	25.4	452	2	US-09-914-259-34	Sequence 34, Appl
31	227	24.9	405	2	US-09-312-283C-374	Sequence 374, Appl
32	224.5	24.6	481	2	US-09-914-259-36	Sequence 36, Appl
33	214.5	23.5	486	2	US-09-914-259-35	Sequence 35, Appl
34	214.5	23.5	486	2	US-09-976-594-278	Sequence 278, Appl
35	214.5	23.5	486	2	US-09-949-016-6216	Sequence 6216, Ap
36	214.5	23.5	507	2	US-09-949-016-9878	Sequence 9878, Ap
37	198.5	21.7	171	2	US-09-996-611D-2	Sequence 2, Appl
38	198.5	21.7	954	2	US-09-996-611D-1	Sequence 1, Appl
39	194	21.2	214	2	US-10-061-658-6	Sequence 6, Appl
40	194	21.2	214	2	US-10-061-658-9	Sequence 9, Appl
41	194	21.2	550	2	US-09-907-794A-227	Sequence 227, App
42	194	21.2	550	2	US-09-905-125A-227	Sequence 227, App
43	194	21.2	550	2	US-09-902-775A-227	Sequence 227, App
44	194	21.2	550	2	US-09-906-700-227	Sequence 227, App
45	194	21.2	550	2	US-09-579-288-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-09-513-999C-6277
Sequence 6277, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duchet, A.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59, US2, REG
CURRENT APPLICATION NUMBER: US/09/513, 999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 6277
LENGTH: 77
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: 66
OTHER INFORMATION: Xaa=Cys or Gly
US-09-513-999C-6277
Query Match
Best Local Similarity 98.7% Pred. No. 4.8e-36;
Matches 76; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 77 MGDHTGALVYAKQLFAEASGARPGVVKLVWVTDGSSDPVPPMOELKDLGVTFI 136
DB 1 MGDHTGALVYAKQLFAEASGARPGVVKLVWVTDGSSDPVPPMOELKDLGVTFI 60
QY 137 VSTGRNPLELSAASA 153
DB 61 VSTGRNPLELSAASA 77
RESULT 2
US-09-949-016-10340
Sequence 10340, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO010307
CURRENT APPLICATION NUMBER: US/09/949,016
```

```

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10340
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10340

Query Match
Best Local Similarity 37.2%; Score 259; DB 2; Length 584;
Matches 67; Conservative 25; Mismatches 84; Indels 4; Gaps 2;
```

```

; 3 DLMFLDSSASVSHYFSRVRREFVQVAPLPLGTGALRASLVHVSRRPYTEFPFGQHS 62
; 348 DLVLVDSKSVRPQNFELVKRFVNIQVDFLDVSEGTGRLVGVSSRVRTFFPLGRYGT 407
; 63 GEAADAVRASQNRGDTHTGTALVYAEQLFAEASGARP---GVPKVLVWVTTDGGSSDP 119
; 408 AAEVQAVALVAYEMRGTMGTGLALRMVHESFSEAQGARPRALNVPVGLVFTDGRSDD 467
; 120 VGPPMQLKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGS 178
; 468 ISWAPRAKEBGIVMIYAVGVGAVERALELREIASPELHVSVAIPDFGTMLLENLRGSI 527
```

```

RESULT 3
US-09-949-016-10341
; Sequence 10341, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10341
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10341
```

```

Query Match
Best Local Similarity 37.2%; Score 259; DB 2; Length 584;
Matches 67; Conservative 25; Mismatches 84; Indels 4; Gaps 2;
```

```

; 3 DLMFLDSSASVSHYFSRVRREFVQVAPLPLGTGALRASLVHVSRRPYTEFPFGQHS 62
; 348 DLVLVDSKSVRPQNFELVKRFVNIQVDFLDVSEGTGRLVGVSSRVRTFFPLGRYGT 407
; 63 GEAADAVRASQNRGDTHTGTALVYAEQLFAEASGARP---GVPKVLVWVTTDGGSSDP 119
; 408 AAEVQAVALVAYEMRGTMGTGLALRMVHESFSEAQGARPRALNVPVGLVFTDGRSDD 467
; 120 VGPPMQLKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGS 178
; 468 ISWAPRAKEBGIVMIYAVGVGAVERALELREIASPELHVSVAIPDFGTMLLENLRGSI 527
```

```

RESULT 4
US-09-919-497-57
; Sequence 57, Application US/09919497
; Patent No. 6773883
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 57
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-57

Query Match
Best Local Similarity 32.3%; Score 256.5; DB 2; Length 755;
Matches 53; Conservative 38; Mismatches 72; Indels 1; Gaps 1;
```

```

; 1 RGDLMFLDSSASVSHYFSRVRREFVQVAPL-PLGTGALRASLVHVSRRPYTEFPFGQ 59
; 5 KADLVFWVDGWSIGDENFNKISFLYSTVGALNKIGDTQVAMVQFTDDPRTDFKUNA 64
; 60 HSGEADAVRASQNRGDTHTGTALVYAEQLFAEASGARPVPKVLVWVTTDGGSSDP 119
; 120 VGPPMQLKDLGVTVFIVSTGRGNFLELSAASAPAEKHLHFVD 163
; 125 VNKISREMQDQYGSIFAGVADADYSELVSGSRARHVFVD 168
```

```

RESULT 5
US-08-897-443-1
; Sequence 1, Application US/08897443
; Patent No. 5981263
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Kaser, Mathew
; TITLE OF INVENTION: HUMAN MATRILIN-3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,443
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0348 US
; TELECOMMUNICATION INFORMATION:
```

GENERAL INFORMATION:

1. APPLICANT: Genentech, Inc.

2. APPLICANT: Ashkenazi, Avi

3. APPLICANT: Borstein, David

4. APPLICANT: Desnoyers, Luc

5. APPLICANT: Eaton, Dan L.

6. APPLICANT: Ferrara, Napoleone

7. APPLICANT: Flivarcoff, Ellen

8. APPLICANT: Fong, Sherman

9. APPLICANT: Gao, Wei-Qiang

10. APPLICANT: Gerber, Hanspeter

11. APPLICANT: Gottlieb, Mary E.

12. APPLICANT: Goddard, A.

13. APPLICANT: Godowski, Paul J.

14. APPLICANT: Grimaldi, Christopher J.

15. APPLICANT: Gurney, Austin L.

16. APPLICANT: Hillan, Kenneth, J.

17. APPLICANT: Kljavin, Ivar J.

18. APPLICANT: Mather, Jennie P.

```

; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905/125A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-125A-34

Query Match      27.5%; Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;
Matches 58; Conservative 36; Mismatches 79;

QY 1 RGDMLFLDSSASVHSEFRREVRVQLVAPLPLGTGLASLVHVSRRPTFFPQGH 60
DB 55 RADLVFIIDSSSVTHDPAKKEFTVDLPGLDGPDVTRVGLQYGSYVNNERSLKT 114
QY 61 SSGEAAQDAVRASAOORMGDTHTGLALVYAKEQLFAEASGARP---GVPRVLVWVVDGSS 117
DB 115 KRKSEVERAVKVRKRLSTGTMTGLAIQYALNLAIFEABEGRRLRENVTRVIMIVDGRPQ 174
QY 118 DPVGPQGLKDLGVTVFIVSTGRNPLELSAASAPAEKHU---HFVVDVLDLHIIVQE 173
DB 175 DSVAEVAAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVFIVANFSQILETLTSVQK 234

RESULT 8
US-09-902-775A-34
; Sequence 34, Application US/09902775A
```

```

; Patent No. 6686451
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Aeshkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kilgavin, Ivar J.
; APPLICANT: Macher, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,775A
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-775A-34
```

```

1  PRIOR FILING DATE: 1999-11-30
2  PRIOR APPLICATION NUMBER: PCT/US99/28564
3  PRIOR FILING DATE: 1999-12-02
4  PRIOR APPLICATION NUMBER: PCT/US99/28565
5  PRIOR FILING DATE: 1999-12-02
6  PRIOR APPLICATION NUMBER: PCT/US99/30095
7  PRIOR FILING DATE: 1999-12-16
8  PRIOR APPLICATION NUMBER: PCT/US99/30911
9  PRIOR FILING DATE: 1999-12-20
10 PRIOR APPLICATION NUMBER: PCT/US99/30999
11 PRIOR FILING DATE: 1999-12-20
12 PRIOR APPLICATION NUMBER: PCT/US00/00219
13 PRIOR FILING DATE: 2000-01-05
14 NUMBER OF SEQ ID NOS: 423
15 SEQ ID NO 34
16 LENGTH: 915
17 TYPE: PRT
18 ORGANISM: Homo sapiens
19 US-09-906-700-34
20
21 Query Match 27.5%, Score 251.5; DB 2; Length 915;
22 Best Local Similarity 32.2%; Pred. No. 2e-19;
23 Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
```

```

; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-603A-34
```

```

Query Match      27.5%; Score 251.5; DB 2; Length 915;
Beet Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

Qy 1 RGDLMFLDSSASVSHVEFSVRREVQGVAPLPFGTGLRSLVHGSRPYTEPFGQH 60
Db 55 RADLVFIIDSSRSVTHYAKKEFIVDLPGLDGPVTRVGLQYGSYVKNESLKTFF 114
Qy 61 SSGEAAQDAVRASAQRMCDTHGTALVYAKQLPAEASGARP---GVPKVLVWYTDGSS 117
Db 115 KRKSEVERAVKMRHLSTGTMTGLAIQVALNTAFSEAEAGARPLENVPVIMIVTDGRPQ 174
Qy 118 DPVGPPMOELKDLGVTTFIVSTGRGNFLELSAAASAPAEKHL-----HFVDVDDLHIYQE 173
Db 175 DSVAEVAARQDITGLIFATIGVGQVDFNTLKSIGSEPHEDHVLVANSQIETLTYSVQK 234
```

```

RESULT 11
US-09-904-920A-34
; Sequence 34, Application US/09904920A
; Patent No. 6806352
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geider, Hanspeter
; APPLICANT: Gerritsen, Maty E.
```

```

; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Macher, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,920A
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-904-920A-34
```

```

Query Match      27.5%; Score 251.5; DB 2; Length 915;
Beet Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

Qy 1 RGDLMFLDSSASVSHVEFSVRREVQGVAPLPFGTGLRSLVHGSRPYTEPFGQH 60
Db 55 RADLVFIIDSSRSVTHYAKKEFIVDLPGLDGPVTRVGLQYGSYVKNESLKTFF 114
Qy 61 SSGEAAQDAVRASAQRMCDTHGTALVYAKQLPAEASGARP---GVPKVLVWYTDGSS 117
Db 115 KRKSEVERAVKMRHLSTGTMTGLAIQVALNTAFSEAEAGARPLENVPVIMIVTDGRPQ 174
Qy 118 DPVGPPMOELKDLGVTTFIVSTGRGNFLELSAAASAPAEKHL-----HFVDVDDLHIYQE 173
```

Db. 175 DSAVAAKARDTGLIFALIGVQVDFTLTKSIGSEPHEDHVLVANSFQIETLTSVFQK 234

RESULT 12
US-09-909-064-34

; Sequence 34, Application US/09909064
; Patent No. 6818449
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,064
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05

; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-064-34

Query Match 27.5%; Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19;
Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

QY 1 RGDLMFLDSSASVSHYEFSEKRVFVGQVAPLPFGTALRSALVHSGRPTFRPFQGH 60
Db 55 RADLVFIIDSSRSVTHDYAKKEPIVDILOFLDGPVTRVGLIQGSTVKNESLKT 114
QY 61 SSGEAAQDAVRASAGRMGDTHGLALVYAKEQLPAEASGARP---GVPELVWMTDGS 117
Db 115 KRKSEVERAKMRHLSTGTMTGLAIQYALNARSSEAGARPLRENVPRVITVDGRQ 174
QY 118 DVGPPMOELKDGLGVTFIVSTGRGNPLELSAASAPAEKHL---HFVDVDLHIYGE 173
Db 175 DSAVAAKARDTGLIFALIGVQVDFTLTKSIGSEPHEDHVLVANSFQIETLTSVFQK 234

RESULT 13

US-09-905-381A-34
; Sequence 34, Application US/09905381A
; Patent No. 6818746

; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,381A
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15

```

; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-381a-34

```

Query Match 27.5%; Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;

Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

```

QY 1 RGDLMFLDSSASVSHYEFVRREFVQGLVAPLPLGTGALRASLVHVGSRPYTEFPFGQH 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 55 RADLVFIIDSSRSVTHYDAKVEFIVDILQFLDIGPVTGVGLQYSGTVKNESSLKTF 114
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 SSGEAAQDAVRASAQRMGDTHTGLALVYAKQPLFAASGARP---GVPRVLVWVTGGSS 117
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 115 KRKSEVERAVKRMRLSTGTMTGLAIQYALNIAFSEAGARPLRENVPRVIMIVDGRQ 174
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 118 DVGPPMOELKDVGTVTVSTGRGNFLELSAASAPAEKHL---HFVDVDDLHIYOE 173
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 175 DSVAEVAAKARDTGILIFAIQGVQVDFNTLKSIGSEPHEDHVFIVANFSQIETLTSVQK 234
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

RESULT 14

US-09-906-618-34

; Sequence 34, Application US/09906618

; Patent No. 6828146

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.

; APPLICANT: Ashkenazi, Avi

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan L.

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerber, Hanspeter

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, A.

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, Christopher J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth, J.

; APPLICANT: Kiljavin, Ivar J.

; APPLICANT: Mather, Jennie P.

; APPLICANT: Pan, James

; APPLICANT: Paoni, Nicholas F.

; APPLICANT: Roy, Margaret Ann

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Williams, P. Mickey

; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,618
; CURRENT FILING DATE: 2001-07-15
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-906-618-34

Query Match 27.5%; Score 251.5; DB 2; Length 915;
Best Local Similarity 32.2%; Pred. No. 2e-19; Indels 7; Gaps 2;

Matches 58; Conservative 36; Mismatches 79; Indels 7; Gaps 2;

```

QY 1 RGDLMFLDSSASVSHYEFVRREFVQGLVAPLPLGTGALRASLVHVGSRPYTEFPFGQH 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 55 RADLVFIIDSSRSVTHYDAKVEFIVDILQFLDIGPVTGVGLQYSGTVKNESSLKTF 114
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 SSGEAAQDAVRASAQRMGDTHTGLALVYAKQPLFAASGARP---GVPRVLVWVTGGSS 117
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 115 KRKSEVERAVKRMRLSTGTMTGLAIQYALNIAFSEAGARPLRENVPRVIMIVDGRQ 174
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 118 DVGPPMOELKDVGTVTVSTGRGNFLELSAASAPAEKHL---HFVDVDDLHIYOE 173
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 175 DSVAEVAAKARDTGILIFAIQGVQVDFNTLKSIGSEPHEDHVFIVANFSQIETLTSVQK 234
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

RESULT 15

US-09-906-646-34

; Sequence 34, Application US/09906646

; Patent No. 6852848

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.

; APPLICANT: Ashkenazi, Avi

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

```

1 APPLICANT: Eaton, Dan L.
2 APPLICANT: Ferrara, Napoleone
3 APPLICANT: Filvaroff, Ellen
4 APPLICANT: Fong, Sherman
5 APPLICANT: Gao, Wei-Qiang
6 APPLICANT: Gerber, Hanspeter
7 APPLICANT: Gerritsen, Mary E.
8 APPLICANT: Goddard, A.
9 APPLICANT: Godowski, Paul J.
10 APPLICANT: Grimaldi, Christopher J.
11 APPLICANT: Gurney, Austin L.
12 APPLICANT: Hillan, Kenneth J.
13 APPLICANT: Kijavini, Ivar J.
14 APPLICANT: Mather, Jennie P.
15 APPLICANT: Pan, James
16 APPLICANT: Paoni, Nicholas F.
17 APPLICANT: Roy, Margaret Ann
18 APPLICANT: Stewart, Timothy A.
19 APPLICANT: Tumas, Daniel
20 APPLICANT: Williams, P. Mickey
21 APPLICANT: Wood, William, I.
22 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
23 TITLE OR INVENTION: Acids Encoding the Same
24 FILE REFERENCE: 10466-14
25 CURRENT APPLICATION NUMBER: US/09/906,646
26 CURRENT FILING DATE: 2002-01-22
27 PRIOR APPLICATION NUMBER: PCT/US00/04414
28 PRIOR FILING DATE: 2000-02-22
29 PRIOR APPLICATION NUMBER: US 60/143,048
30 PRIOR FILING DATE: 1999-07-07
31 PRIOR APPLICATION NUMBER: US 60/145,698
32 PRIOR FILING DATE: 1999-07-26
33 PRIOR APPLICATION NUMBER: US 60/146,222
34 PRIOR FILING DATE: 1999-07-28
35 PRIOR APPLICATION NUMBER: PCT/US99/20594
36 PRIOR FILING DATE: 1999-09-08
37 PRIOR APPLICATION NUMBER: PCT/US99/20944
38 PRIOR FILING DATE: 1999-09-13
39 PRIOR APPLICATION NUMBER: PCT/US99/21090
40 PRIOR FILING DATE: 1999-09-15
41 PRIOR APPLICATION NUMBER: PCT/US99/21547
42 PRIOR FILING DATE: 1999-09-15
43 PRIOR APPLICATION NUMBER: PCT/US99/23089
44 PRIOR FILING DATE: 1999-10-05
45 PRIOR APPLICATION NUMBER: PCT/US99/28214
46 PRIOR FILING DATE: 1999-11-29
47 PRIOR APPLICATION NUMBER: PCT/US99/28313
48 PRIOR FILING DATE: 1999-11-30
49 PRIOR APPLICATION NUMBER: PCT/US99/28564
50 PRIOR FILING DATE: 1999-12-02
51 PRIOR APPLICATION NUMBER: PCT/US99/28565
52 PRIOR FILING DATE: 1999-12-02
53 PRIOR APPLICATION NUMBER: PCT/US99/30095
54 PRIOR FILING DATE: 1999-12-16
55 PRIOR APPLICATION NUMBER: PCT/US99/30911
56 PRIOR FILING DATE: 1999-12-20
57 PRIOR APPLICATION NUMBER: PCT/US99/30999
58 PRIOR FILING DATE: 1999-12-20
59 PRIOR APPLICATION NUMBER: PCT/US00/00219
60 PRIOR FILING DATE: 2000-01-05
61 NUMBER OF SEQ ID NOS: 423
62 SEQ ID NO 34
63 LENGTH: 915
64 TYPE: PRT
65 ORGANISM: Homo sapiens
66 US-09-906-646-34

```

Query Match	27.5%	Score 251.5	DB 2	Length 915
Best Local Similarity	32.2%	Pred. No. 2e-19		
Matches 58	36	Mismatches 79	Indels 7	Gaps 2

QY 1 RGDMLPDDSSASVSHYEFSSRVAFVQGLVAPLPDGTGALRASLVHGVSSPPTYEEFGQH 60

```

Db      55 RADLVPIIDSSSYNHDYAKYKEFIYDILQFLDIGDVTFRVGLQGSYTKANEFSLKT 114
OY      61 SSGEAAQDVAVRSAQRMGDTHGLALVTYAKQLFPAEASGARP---GYPKVLVWYTDGSS 117
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      115 KRKSEVERAVKRMHSLSTGTMTGLIYQVALINAFSEAGAPLRRENPVRVIMITVDRPQ 174
OY      118 DPVGPPEMQLKDLGVTVFIVSTGRGNPLELSAASAPAEKGL-----HFVVDLHIIYOE 173
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      175 DSYAEVAARAKRDGILIFAIVGQVDFTLKSIGSEPEHEDVFLVNFNSQIETLTSVFOK 234

```

Search completed: February 13, 2006, 07:50:51
Job time : 28.6923 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:40:09 ; Search time 62.2107 Seconds
(without alignments)
646,490 Million cell updates/sec

Title: US-10-699-035A-6

Perfect score: 2154

Sequence: 1 MPMWALGALSLRLALARS.....RRPRPVPRAPTGTASREP 418

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437.5	20.3	3124	2 A40020	collagen alpha 1(X)
2	382.5	17.8	2944	2 A54849	collagen alpha 1(V)
3	380.5	17.7	1857	2 S31212	collagen alpha 1(X)
4	380.5	17.7	1888	2 S78476	collagen alpha 1(X)
5	369.5	17.2	1747	2 A45974	collagen alpha 1(X)
6	278	12.9	929	2 I51027	type XII collagen
7	268.5	12.5	843	2 A40970	undulin 1 - human
8	248.5	11.5	493	2 A33809	cartilage matrix p
9	243.5	11.3	500	2 S66522	cartilage matrix p
10	239	11.1	3137	2 A37797	collagen alpha 3(V)
11	233.5	10.8	496	2 A37979	cartilage matrix p
12	225.5	10.5	3176	2 CGHJ3A	collagen alpha 3(V)
13	220	10.2	741	2 T46488	hypochelical prote
14	209	9.7	272	2 A55348	integrin alpha-1-c
15	208	9.7	1151	2 A45226	integrin alpha-1-c
16	202	9.4	1180	2 A35854	protein alpha-1-c
17	183	8.5	1286	2 A48396	protein M01E10.2
18	181.5	8.4	724	2 A48569	antigen Em100 - E1
19	179	8.3	550	2 T23760	hypochelical prote
20	174.5	8.1	2481	2 A43908	fibronectin - Afri
21	174	8.1	1153	1 RWHUB	cell surface glyco
22	173	8.0	1153	1 S00551	leukocyte surface
23	170.5	7.9	1810	1 A32230	tenascin precursor
24	169.5	7.9	3051	1 S42373	hypochelical prote
25	168.5	7.8	1163	1 RWHUB	cell surface glyco
26	168	7.8	712	1 A45618	immunodominant mic
27	167	7.8	427	2 G00039	von Willebrand fac
28	166.5	7.7	1746	1 S19694	tenascin precursor
29	165.5	7.7	2201	2 A32160	tenascin-C - human

30	165	7.7	1163	2 I56126	lymphocyte function
31	164	7.6	3566	1 A40701	tenascin-X precurs
32	163.5	7.6	1170	2 I45914	integrin alpha 2 s
33	163	7.6	4135	2 T42629	tenascin-X - bovin
34	162.5	7.5	933	2 A31930	cytotactin - chick
35	162.5	7.5	2386	1 FNHU	fibronectin precur
36	162	7.5	2813	1 VWHU	von Willebrand fac
37	161	7.5	1178	2 S44142	VLA-2 protein homo
38	160	7.4	1907	2 S50893	protein-tyrosine-P
39	157.5	7.3	1020	2 A29355	fibronectin - chic
40	154	7.1	2265	1 FNBO	fibronectin - bovi
41	153.5	7.1	2019	1 J03322	tenascin precursor
42	152	7.1	2477	2 S14428	fibronectin precur
43	151.5	7.0	1807	2 J06319	integrin beta-4 ch
44	151	7.0	1914	2 T42635	tenascin Y precurs
45	150.5	7.0	1170	2 S03308	cell surface glyco

ALIGNMENTS

RESULT 1
A40020
collagen alpha 1(XII) chain precursor - chicken
N/Alternate names: fibronectin
C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #ext change 09-Jul-2004
R/Accession: A40020; A34485; B34485; A28037; S23814; S22554; S28811
R/Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinmura, T.; Tanaka, H.; Nishida, Y.; Oba
U. Cell Biol. 115, 209-221, 1991
A/Title: The complete primary structure of type XII collagen shows a chimeric molecule
nous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site
A/Reference number: A40020; MWID:92011862; PMID:1918137
A/Accession: A40020
A/Molecule type: mRNA
A/Residues: 1-3124 <YAM>
A/Cross-references: UNIPROT:P13944; UNIPARC:UPI0000126D2D; GB:D00824; NID:G222810; PIDN
A/Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and
R/Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
J. Biol. Chem. 264, 19772-19778, 1989
A/Title: Type XII collagen: A large multidomain molecule with partial homology to type
A/Reference number: A34485; MWID:90062079; PMID:2584192
A/Accession: A34485
A/Molecule type: mRNA
A/Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
A/Cross-references: UNIPARC:UPI0000171233; EMBL:J05137; NID:G211284; PIDN:AAA48635.1; P
A/Accession: B34485
A/Molecule type: protein
A/Residues: 2772-2792; 2846-2873 <GOR2>
A/Cross-references: UNIPARC:UPI0000173C3F; UNIPARC:UPI0000173C40
R/Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A/Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA
A/Reference number: A28037; MWID:87317590; PMID:3476525
A/Accession: A28037
A/Molecule type: mRNA
A/Residues: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>
A/Cross-references: UNIPARC:UPI0000171224; EMBL:M17375; NID:G211649; PIDN:AAA48718.1; P
A/Note: This sequence has been revised in reference A34485
R/Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992
A/Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A/Reference number: S23814; MWID:92362621; PMID:1323460
A/Accession: S23814
A/Molecule type: protein
A/Residues: 'X', 1333, 'Q', 1335-1347, 1914-1928, 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517
A/Cross-references: UNIPARC:UPI0000173C41; UNIPARC:UPI0000173C42; UNIPARC:UPI0000173C43
R/Dublet, B.; van der Rest, M.
J. Biol. Chem. 262, 17724-17727, 1987
A/Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of peptin
A/Reference number: S22254; MWID:88087065; PMID:3121603
A/Accession: S22254
A/Molecule type: protein

A:Residues: 2831-2832, 'T', 2834, 'R', 2836-2843, 3002-3014 <DOB>
A:Cross-references: UNIPARC:UPI0000173C44; UNIPARC:UPI0000173C45
R:Trueb, U.; Trueb, B. Biochim. Biophys. Acta 1171, 97-98, 1992
A:Title: The two splice variants of collagen XII share a common 5' end.
A:Accession number: S28811; MUID:93042014; PMID:1420368
A:Reference: S28811
A:Accession: S28811
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>
A:Cross-references: UNIPARC:UPI0000173C46; EMBL:X67327
C:Genetics:
A:Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3065/1
C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disulfide
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
F:24, 1189-1124/Product: collagen alpha 1(XII) chain short splice form #status predicted
F:24-114/Domain: IIIA #status predicted <IIIA>
F:24-105/Domain: fibronectin type III repeat homology <FN3A>
F:137-301/Domain: von Willebrand factor type A repeat homology <VWA1>
F:332-425/Domain: IIB #status predicted <IIB>
F:332-414/Domain: fibronectin type III repeat homology <FN3B>
F:437-601/Domain: von Willebrand factor type A repeat homology <VWA2>
F:629-1178/Domain: IIIC #status predicted <IIIC>
F:7630-711/Domain: fibronectin type III repeat homology <FN3C>
F:721-802/Domain: fibronectin type III repeat homology <FN3D>
F:812-885/Domain: fibronectin type III repeat homology <FN3E>
F:905-986/Domain: fibronectin type III repeat homology <FN3F>
F:995-1076/Domain: fibronectin type III repeat homology <FN3G>
F:1086-1169/Domain: fibronectin type III repeat homology <FN3H>
F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>
F:1384-1465/Domain: IIID #status predicted <IIID>
F:1384-1457/Domain: fibronectin type III repeat homology <FN3I>
F:1474-1557/Domain: fibronectin type III repeat homology <FN3J>
F:1586-1647/Domain: fibronectin type III repeat homology <FN3K>
F:1655-1738/Domain: fibronectin type III repeat homology <FN3L>
F:1756-1838/Domain: fibronectin type III repeat homology <FN3M>
F:1847-1928/Domain: fibronectin type III repeat homology <FN3N>
F:1937-2019/Domain: fibronectin type III repeat homology <FN3O>
F:2028-2110/Domain: fibronectin type III repeat homology <FN3P>
F:2119-2199/Domain: fibronectin type III repeat homology <FN3Q>
F:2207-2294/Domain: fibronectin type III repeat homology <FN3R>
F:2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>
F:2438-2440/Region: cell adhesion #status predicted
F:2509-3750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted
F:2751-2902/Domain: collagenous COL2 #status predicted <COL2>
F:2899-2901/Region: cell attachment (R-G-D) motif
F:2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>
F:2946-3024/Domain: collagenous COL1 #status predicted <COL1>
F:3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>
F:32, 1006-1032, 1044-1512, 1767, 2210, 2273, 2532, 2683/Binding site: carboxylate (Asn) (covar
F:2780, 2789, 2836, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) #stat

[illegible]

A;Residues: 2395-2871,'S',2873-2944 <RE2>
A;Cross-references: UNIPARC:UPI00001A724; GB:L06862; NID:G388713; PIDN:AAA89196.1; PID:
R;Christiano, A.M.; Ryyanen, M.; Uitto, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A;Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --- Ser substitution
A;Reference number: A5555; MIMD:94224777; PMID:8170945
A;Content: annotation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL7A1; EBR1; EBD1; EB
A;Cross-references: GDB:128750; OMIM:120120
A;Map position: 3p21.3-3p21.3
A;Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis
A;Note: there are 118 introns
C;Complex: type VII collagen is probably a homotrimer
C;Function:
A;Description: structural component of extracellular polymer associated with anchoring
F;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F;17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F;36-201/Domain: von Willebrand factor type A repeat homology <VMA1>
F;331-318/Domain: fibronectin type III repeat homology <FN1>
F;327-313/Domain: fibronectin type III repeat homology <FN2>
F;414-502/Domain: fibronectin type III repeat homology <FN3>
F;508-593/Domain: fibronectin type III repeat homology <FN4>
F;598-683/Domain: fibronectin type III repeat homology <FN5>
F;686-771/Domain: fibronectin type III repeat homology <FN6>
F;776-862/Domain: fibronectin type III repeat homology <FN7>
F;864-952/Domain: fibronectin type III repeat homology <FN8>
F;954-1045/Domain: fibronectin type III repeat homology <FN9>
F;1052-1219/Domain: von Willebrand factor type A repeat homology <VMA2>
F;1170-1172/Region: cell attachment (R-G-D) motif
F;1189-1253/Region: cysteine/proline-rich
F;1234-2783/Region: interrupted helical
F;1334-1336/Region: cell attachment (R-G-D) motif
F;2008-2010/Region: cell attachment (R-G-D) motif
F;2553-2555/Region: cell attachment (R-G-D) motif
F;2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F;2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <NP1>
F;337,786,1109/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
F;2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F;2625,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;2634,2802,2804/Disulfide bonds: interchain #status predicted

Query Match 17.8%; Score 382.5; DB 2; Length 2944;
Best Local Similarity 31.0%; Pred. No. 1.2e-18;
Matches 136; Conservative 46; Mismatches 208; Indels 49; Gaps 11;

11 LSLRLALRSGARGPASAPR-----GDMFLDSSASVSHYEFRRV 54
1 MTRRLVLAALCA--GILAEAPRPAQHRERVTCTRLVADIVFLDSSSSIGSNFEVR 58
55 EFVGVLAAPLPLGTGA--LRASLVHVSRRPYTEFPFGHSSGEAODAVASQRMGDT 112
59 SFLEGLVLPFGSAGASQGVAFATVQYSDDRTEFGDALGSGGVIRAIHELKYGKGNTR 118
113 TGLALVYAEQQLFAEASGAPGVKLVWYTDGSSDPVPPMOELKDLGVTVFYVSTGR 172
119 TGAALIHVADHVELPOL-ARPGVAKVCILITDGKSQDLVDTAAQLKGGGVKFAVGIKN 177
173 GNLELELSAASAPAEKRLHVDVDDLHII-----VQELRGSILDMMP 215
178 ADBELKRVASQPTSDFFP--VNDPSILRTLPLVSRVCTTAGVPTVRPPDDTSAP 235
216 QQLHATEITSSGFLAMPPLLTADSGYV--LLEIVPSAOPGAARQO--LPGNATDWIWA 271
236 RDLVLSPPSSQSLRVQWTAASGPVTVGKYQVTLPLTGQGLPBERGQVNVVPAGETSVRLR 295
272 GLDPDIDVAVLPESNVRLRLPOLLRVTRPREBAGERIVISHARPSLRVSWAPALGS 331

296 GLRPLETVQVTVIALYANSIGENVGSTATTALF-GPE-LTIOHTAHSILVWRSVPG- 352
332 AALGYHVOFGPLRGGAORVEVPAGRNCTTGLAPGATVYLVTAAPRSRESALSAR 391
353 --ATGRTVWVRVLSGGPFTQQELGPGQSVLRDLRPGTDVETVTVTLFGRSVGPATSLM 410
392 ACTPDGPRPPRPVPPAPT 410
411 ARTDASVEQLRPLVILGPT 429

RESULT 3
S31212
collagen alpha 1(XIV) chain precursor, short form - chicken
C;Species: Gallus gallus (chicken)
C;Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #ext_change 15-Sep-2003
C;Accession: S31212
R;Welch, C.; Trub, J.; Kessler, B.; Winterhalter, K.H.; Trub, B.
Eur. J. Biochem. 212, 483-490, 1993
A;Title: Complete primary structure of chicken collagen XIV.
A;Reference number: S31211; MIMD:93185668; PMID:8444186
A;Accession: S31212
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-1857 <MA5>
A;Cross-references: UNIPARC:UPI000017123C; EMBL:X70792; NID:G288874; PIDN:CAA50063.1; P
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Genetics:
A;Gene: COL14A1
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>
F;29-110/Domain: fibronectin type III repeat homology <FN3A>
F;156-320/Domain: von Willebrand factor type A repeat homology <VMA1>
F;352-433/Domain: fibronectin type III repeat homology <FN3B>
F;442-525/Domain: fibronectin type III repeat homology <FN3C>
F;534-614/Domain: fibronectin type III repeat homology <FN3D>
F;623-707/Domain: fibronectin type III repeat homology <FN3E>
F;741-823/Domain: fibronectin type III repeat homology <FN3F>
F;832-914/Domain: fibronectin type III repeat homology <FN3G>
F;922-1009/Domain: fibronectin type III repeat homology <FN3H>
F;1040-1205/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 17.7%; Score 380.5; DB 2; Length 1857;
Best Local Similarity 28.6%; Pred. No. 9.1e-19;
Matches 117; Conservative 63; Mismatches 158; Indels 71; Gaps 13;

27 PASAPRGDLMFLDSSASVSHYEFRRVREFVGVLAAPLPLGTALASLVHVSRRPYTER 86
154 PAIA--DIVILVDGWSIGRFNFRVLRFLLENVSAFVNGSEKTVGLAQYSGDPRIEW 210
211 HUNAATGKDVLDVAVNNLPYKGSNTLTGLALTYILENSFKPEAPGAPGSKIGILTIDK 270
147 SSDPVPMPMOELKDLGVTVFYVSTGRGNLELSAASAPAEKRLHVV-DVDDLHIIVQ-- 203
271 SQDDVLPAAKMLRDAGISLPAIGVNADINELKEIASBPSTHVNVVADPNFMSIYEG 330
204 -----ELRGSILDM-RPQQLHATEITSSGFLAMPPLLTADSG----- 241
331 TRTVCSRVREOEKEIGTIAASLGAPTDLVTSITARGRVSV-----THSPGVEXRRVV 386
242 YVVLLEIVPSAOPGAARQOQLPGNATWIMAGLDPDIDVAVL-----PESN 288
387 YV-----PFGQGPBEVVVDSSSTAVLKNLMSLTYQIAYVAITSNAASEGLRTETT 440
289 VRLRLPOLLRVTRPREBAGERIVISHARPSLRVSWAPALGSAAALGYHVOGFLRG- 347
441 LAL-----PMAADLKLYDVSHS---SBRAKMN---GVAAGATGVWILVLAAPLDEGL 483
348 EAQREVPVPRNCTTQ--GLAPGATVYLVTAAPRSRESALSARACT 394

Db 484 AADEKEIKIGEASTELDLGLPNTYVTVYVAMGEASDPLTGQETT 532

RESULT 4
S78476
collagen alpha 1(XIV) chain precursor, long form - chicken
C/Species: Gallus gallus (chicken)
C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 09-Jul-2004
C/Accession: S78476; S31211
R./Trueb, B.
Submitted to the EMBL Data Library, January 1993
A/Reference number: S78476
A/Accession: S78476
A/Molecule type: mRNA
A/Residues: 1-1888 <TRU>
A/Cross-references: UNIPROT:P32018; UNIPARC:UPI0000126D31; EMBL:X70793; NID:G288872; PID:R.Maechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A/Title: Complete primary structure of chicken collagen XIV.
A/Reference number: S31211; MUID:93185668; PMID:844186
A/Accession: S31211
A/Status: preliminary
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-416; 1460-1811, 1843-1888 <MAE>
A/Cross-references: UNIPARC:UPI00001773F4; UNIPARC:UPI00001773F5; EMBL:X70793
C/Genetics:
A/Gene: COL14A1
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F.1-28/Domain: signal sequence #status predicted <SIG>
F.29-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <MAT>
F.29-110/Domain: fibronectin type III repeat homology <FN3A>
F.156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F.352-433/Domain: fibronectin type III repeat homology <FN3C>
F.442-525/Domain: fibronectin type III repeat homology <FN3C>
F.534-614/Domain: fibronectin type III repeat homology <FN3C>
F.623-707/Domain: fibronectin type III repeat homology <FN3C>
F.741-823/Domain: fibronectin type III repeat homology <FN3C>
F.832-914/Domain: fibronectin type III repeat homology <FN3C>
F.922-1009/Domain: fibronectin type III repeat homology <FN3C>
F.1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 17.7%; Score 380.5; DB 2; Length 1888;
Best Local Similarity 28.6%; Pred. No. 9.3e-19;
Matches 117; Conservative 63; Mismatches 158; Indels 71; Gaps 13;

Qy 27 PASAPRGDMLPLDSSASVSHYFSRVRFEVQGLVAPLPGTALPASLVHGSRPYTEF 86
Db 154 PAIA---DIVILVDGWSIGRNFRLVRLFLNLVSAFVSGSKTRVGLAQYSGDPRTEM 210
Qy 87 PFGQSSGEAODAVRASAQRMGDTHGLALVYAKSQLFAEASGARPGVPKYLWVTTDGG 146
Db 211 HLNAGYTKDAVLDVAVNLPRYKGGNTLTGLALFTYILENSFKPEAGAPVSKIGILTDDK 270
Qy 147 SSDPVGPRMQLKDLGVTFYVSTGRNLELSAASAAPAEKILHVV-DVDDLIHIVQ-- 203
Db 271 SDDDVTPPAKNLRDAIEIFAIGVKNADINEKEIASBDSITHVVAADFNFNSIVEGL 330
Qy 204 -----ELRGSILDM-PPQQLHATEITSSGRLAPPLTADSG----- 241
Db 331 TRTVCSRVEEKEIKIGASTELDLGLPNTYVTVYVAMGEASDPLTGQETT 386
Qy 242 YVLELVPSAQPGAAARQQLPGNATDWIAGLDPTDIDVALV-----PESN 288
Db 387 YV-----PTRGGQPREVVVDGSSPAVLKNLMSLTREYQIVAFIYSNASEGLRGTEET 440
Qy 289 VLLLRPQILIRKTRPEEAPERIVISHAPRSIRVSWAALSGAALGTHVQPGPLRG- 347
Db 441 IAL-----PMSADLKLYDVSHS---SMRAKN---GVAAGTGVMILYAPLDEGL 483
Qy 348 EAQRVVEPAGRNCTLO--GLAPGTAYLVTVTAAPFSGRESALSAACT 394
Db 484 AADEKEIKIGEASTELDLGLPNTYVTVYVAMGEASDPLTGQETT 532

RESULT 5
A45974
collagen alpha 1(XIV) chain precursor, short form 2 - chicken
N/Alternate names: undulin
C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: A45974; S30085; S22916; S17035; S20833
J.Giercke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.; Lin
J. Biol. Chem. 268, 12177-12184, 1993
A/Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region
ns
A/Reference number: A45974; MUID:93280195; PMID:8505337
A/Accession: A45974
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-1747 <GER>
A/Cross-references: UNIPROT:P32018; UNIPARC:UPI0000173C47
A/Experimental source: embryo skin
A/Note: sequence inconsistent with the nucleotide translation
R./Apte, S.S.
Submitted to the EMBL Data Library, March 1992
A/Reference number: S30085
A/Accession: S30085
A/Molecule type: mRNA
A/Residues: 1472-1660 <APR>
A/Cross-references: UNIPARC:UPI0000006A2; EMBL:X65122; NID:G62871; PID:CAA46238.1; PID:R./Trueb, J.; Trueb, B.
Eur. J. Biochem. 207, 549-557, 1992
A/Title: Type XIV collagen is a variant of undulin.
A/Reference number: S22916; MUID:92339443; PMID:1339349
A/Accession: S22916
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 286-494; 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>
A/Cross-references: UNIPARC:UPI0000173C48
R./Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsemayer, T.F.; van der Rest, M.; Mayne, Eur. J. Biochem. 201, 333-338, 1991
A/Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens
A/Reference number: S17035; MUID:92037585; PMID:1935930
A/Accession: S17035
A/Molecule type: mRNA
A/Residues: 1472-1659 <GOR1>
A/Cross-references: UNIPARC:UPI0000173C49
A/Accession: S20833
A/Molecule type: protein
A/Residues: 1551-1570; 1593-1639-1667 <GOR2>
A/Cross-references: UNIPARC:UPI0000173C49
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F.40-204/Domain: von Willebrand factor type A repeat homology <VWA1>
F.236-317/Domain: fibronectin type III repeat homology <FN3A>
F.326-409/Domain: fibronectin type III repeat homology <FN3A>
F.418-498/Domain: fibronectin type III repeat homology <FN3C>
F.507-591/Domain: fibronectin type III repeat homology <FN3D>
F.625-707/Domain: fibronectin type III repeat homology <FN3B>
F.716-798/Domain: fibronectin type III repeat homology <FN3F>
F.806-893/Domain: fibronectin type III repeat homology <FN3G>
F.924-1069/Domain: von Willebrand factor type A repeat homology <VWA2>
F.1111-1152/Domain: non-collagenous NC4 #status predicted <NC4>
F.1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>
F.1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Query Match 17.2%; Score 369.5; DB 2; Length 1747;
Best Local Similarity 28.1%; Pred. No. 5.1e-18;
Matches 115; Conservative 64; Mismatches 159; Indels 71; Gaps 13;

Qy 27 PASAPRGDMLPLDSSASVSHYFSRVRFEVQGLVAPLPGTALPASLVHGSRPYTEF 86
Db 38 PAIA---DIVILVDGWSIGRNFRLVRLFLNLVSAFVSGSKTRVGLAQYSGDPRTEM 94
Qy 87 PFGQSSGEAODAVRASAQRMGDTHGLALVYAKSQLFAEASGARPGVPKYLWVTTDGG 146

```

Db      95  HLNMYGTADALVDADRNLPRYKGNTLTGLATLYILENCFKEBAGARPVSXIGILITDGK 154
QY      147  SSDVGPMPQELKOLGVTVFVSTGRGNFLELSAAASPAKRLHFV-DVVDLHIYQ-- 203
Db      155  SQDDVIPAKRLRAGLELFAIGVKNAIDINLKEIASEPDSITHYVNADEFPMKSIYEG 214
QY      204  -----ELRGSILDM-RPQOLHATEITSSGFRFLAMPPLPADSG----- 241
Db      215  TRTVCSRYBEQEKIKGTIASSLGAFTDLVNSDITARGFRYSW-----THSPGVKRYRV 270
QY      242  YYVELVPSAQPAGARRQQLFGNATDWMAGLDPTDYDALV-----PESN 288
Db      271  YY-----PYRGGPREEVYVDGSSSTAVALKMLMSTEQIVAFIYSNAASEGLRGTETT 324
QY      289  VRLRLPOLRLVTRTPPEEAGPERIVISHARPSRLIVSWAPALGSAALGYHQFGLRGG- 347
Db      325  LAL-----PMASDLKLDVYSHS---SMAKWN---GVAGATGYMLLYAPLTGEL 367
QY      348  EAQGVEPAGRNCTTLO--GLAPGTALVYVTTAAFRSREBALSAAKACT 394
Db      368  AADEKEIKIGEASTLELDGLLPNTETVTVVYAMFGEASDPLTGQETT 416

```

```

RESULT 6
151027
type XII collagen alpha-1 chain - eastern newt (fragment)
C/Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C/Accession: 151027
C/Mel, Y.; Yang, E.V.; Klatte, K.P.; Tassava, R.A.
Dev. Biol. 168 503-513, 1995
A/Title: Monoclonal antibody M72 identifies the urodele alpha 1 chain of type XII colla
A/Reference number: 151027; MUID:95246925; PMID:7729585
A/Accession: 151027
A/Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-929 <WEI>
A:Cross-references: UNIPROT:Q91145; UNIPARC:UPI000012602E; EMBL:U19494; NID:G632647; PIR
F_115-226/Domain: fibronectin type III repeat homology <3FR>
F_631-795/Domain: von Willebrand factor type A repeat homology <VMA3>

```

```

Query March          12.9#: Score 278; DB 2; Length 929;
Best Local Similarity 24.0#: Pred. No. 7.2e-12;
Matches    87; Conservative   51; Mismatches 129; Indels   96; Gaps   7;

QY      22 AERGP-PASA-----PRGDLMFLLDSSASVSHYESFVKREFFGVQLAFLPLTGALRAS 74
           :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      614 APRPFISSGLDCTTKKAQADIVLLVDGSWSIGRPNFKIVRNFIIRVAVEFDIGSDRVQLA 673

QY      75 LVHVGSRPYTEFFPPGGCHSSGGEAAODAVRASKQRNGDTHTGLALYYAKEQLFAEASGARPG 134
           :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      674 VSOYSGDPRTEWQNTHTKTKSLMDAVANLPYKGAGNTWTGALKFTLENNRPGVMPEK 733

QY      135 VPKLVLMWTDGSGSDDPVGPMPWOELKDGLVTAFIVSTGRGNFLEKSAAMASAPAEKHLPV- 193
           :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      734 ARKAIITLTDCKSQDDIIVAPSKRYADESIELIXANGIKRADENELKEAISDPDELYMNVVA 793

QY      194 DVDDLHIITVOELRGSIILDMARPQQLHATEITSSGFRLAMPPELLTAISGYVLELVPSAQP 253
           :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      794 DFSLLTINI VNDLTENVCSYVKG----- 815

QY      254 GAARQQQLPGAATWIMAGLDPDPTDYVALVPESNVLLRPQILLRVTRPEABEPKIVI 313
           :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      816 -----PG-----GLNP-----PSNLVT 827

QY      314 SHARPRLIRVSWAPALGSAALGYHVQGPLRGSEACQREVVPAGRNCITTLLOGLAPGTAYL 373
           :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      828 SEFPSPRRVTVWPP--SQSVRFKVEYEPYVAGGAPQGCVYVRGQTTLVLGLMPETELY 885

QY      374 VTV 376
           :|
Db      886 VNV 888

```

RESULT 7

A40970

undulin 1 - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004

C/Accession: A40970

R/Juett, M.; Herbst, H.; Hummel, M.; Duerkop, H.; Tippler, D.; Stein, H.; Schuppan, D.

J. Biol. Chem. 266, 17366-17373, 1991

A/Title: undulin is a novel member of the fibronectin-tenascin family of extracellular

A/Reference number: A40970; MUID:91373351; PMID:1176629

A/Accession: A40970

A/Molecule type: mRNA

A/Residues: 1-843 <JUS>

A/Cross-references: UNIPROT:Q05707; UNIPARC:UPI00000716A9; GB:M64108; NID:G340081; PIDN

C/keywords: glycoprotein

F/165-246/Domain: fibronectin type III repeat homology <FN3A>

F/255-338/Domain: fibronectin type III repeat homology <FN3B>

F/347-427/Domain: fibronectin type III repeat homology <FN3C>

F/436-520/Domain: fibronectin type III repeat homology <FN3D>

F/547-632/Domain: fibronectin type III repeat homology <FN3E>

F/641-723/Domain: fibronectin type III repeat homology <FN3F>

F/731-818/Domain: fibronectin type III repeat homology <FN3G>

	Query March	12.5%	Score 268.5;	DB 2;	Length 843;
	Best Local Similarity	25.7%	Pred. No. 3e-11;		
	Matches	95;	Conservative	53;	Mismatches 152; Indels 69; Gaps 12;
Qy	66 LGTGLRSLVHVGRSPRYTEPPFGQHSSGEAADAVFASAKRMGDTHHTGLALVYAKEOLF	125	:	:	:
	: :	:	:	:	:
Sb	3 VGSEKRTIGLAOYSGDPRIEMHNAFSTKCEVEIEAVGNLPHYKGNNLTITGLALNFIENFSF	62	:	:	:
Qy	126 ABAAGARPGVKYLWWTGDSGSDPVGPRMOEIKDLGVTVPIVSTGRGNFLSAAASAP	185	:	:	:
Db	63 KPEAEGSTGVSKIGILITTDGKSODDIIIPSRNLRESGEVELFALGVKADANLELOETISER	122	:	:	:
Qy	186 AEKHIAHFV-DVDDLIHIIVOEIRGSILDAMR-----POOLHATEITSSGF	228	:	:	:
Db	123 DSTHYVVVAEFDMHMTVESLTRTLCSRVEEDPREIKASAHATIGRPTELITSEVTARSF	182	:	:	:
Qy	229 RLAMPRLITLDSC-----YYVLIELVPSAQGAARQOLPGNATMIWAGLDPPDDYD	280	:	:	:
Db	183 MVVN---TTAPGVKEKYRVVY---PTRGGKDEVVVDGVSSTLVKNMLSTLEYQ	232	:	:	:
Qy	281 VALVPESNVRLRLPOLLRVTRPEE-AGPER-----IVISHAPRSLSRWSPALG	330	:	:	:
Db	233 IAYF-----AIYAHTASEGLRGCTETTLTPMASDLLLYDVENSMRWKMDAVPG	281	:	:	:
Qy	331 SAALGYHVOFGPLRGSEA-QRVEVPAGRNC-TLOGIAPGTAYLVTVTAAFR-----	381	:	:	:
Db	282 ---ASGYLLIYAPLTEGLADEKEMKIGEHTDIETLSGLLPNTPEYTVTVVAMFGEBSADP	338	:	:	:
Qy	382 -SGRESALS 389		:	:	:
Db	339 VTGOETTILA 347		:	:	:

RESULT 8
A33809
cartilage matrix protein precursor - chicken
C/Species: Gallus gallus (chicken)
C/Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 09-Jul-2004
C/Accession: A33809; A25364
R/Klars, I., Deak, F., Holloway Jr., R.G., Delius, H., Meubst, K.A., Frimberger, E.; Arg
J. Biol. Chem. 264, 8126-8134, 1989
A/Title: Structure of the gene for cartilage matrix protein, a modular protein of the e-
grins, Von Willebrand factor, complement factors B and C2, and epidermal growth factor.
A/Reference number: A33809; MUID:89255246; PMID:2542265
A/Accession: A33809
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-493 <K145>
A/Cross-references: UNIPROT:P05099; UNIPARC:UP1000004F1BF; GB:X12346; GB:X12347; GB:X12

R:Argaves, W.S.; Deak, F.; Sparks, K.J.; Kiss, I.; Goettinck, P.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 464-468, 1987
C>Title: Structural features of cartilage matrix protein deduced from cDNA.
A:Reference number: A26364; MUID:87092429; PMID:3025875
A:Accession: A26364
A:Molecule type: mRNA
A:Residues: 78-493 <ARG>
A:Cross-references: UNIPARC:UPI000004F1EE; GB:M14792; NID:g211545; PIDN:AAA48695.1; PID:
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repeat
F:37-204/Domain: von Willebrand factor type A repeat homology <VMA1>
F:225-260/Domain: EGF homology <EGF>
F:270-434/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 11.5%; Score 248.5; DB 2; Length 493;
Best Local Similarity 37.1%; Pred. No. 4e-10;
Matches 66; Conservative 22; Mismatches 83; Indels 7; Gaps 2;

QY 25 GPPASPRG-----DLMFLDSSASVSHYEFNRVREFVQGLVAPLPLGTGALRASLVHVV 78
DB 257 GRTCSAGSGSGALDVLFLDSSKSVRPENFVLKFTNQIVESLEVBKQAQVGLVQY 316
QY 79 GSRPYTEPFPGQHSGEAAQDAVRASQRMGDHTHTGLALVYAKEQLFAEASGARPGVPVY 138
DB 317 SSSVRGEFPLGGQKNNKDIKAAVKQAAVMEKGMTGQALKYLVDSFSIANGARPGVPVY 376
QY 139 LVMVTDGSSDPVPGPMQELKDIGTVFYVSTGRGNFLISAASAPAEKHLHFVDV 196
DB 377 GIVFTDGRSDYITDAKAKADLGFPMFVAGVNAVEDELRISAEPVAEH-YFYTAD 433

RESULT 9
S66522
Cartilage matrix protein precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S66522
Burr, J. Biochem. 236, 970-977, 1996
A>Title: Cloning, sequencing and expression analysis of mouse cartilage matrix protein c
A:Reference number: S66522; MUID:96270751; PMID:8665920
A:Accession: S66522
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-500 <AS2>
A:Cross-references: UNIPROT:P51942; UNIPARC:UPI0000029480; EMBL:U55035; NID:g1163178; PI
C:Genetics: C
A:Gene: CMP
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-500/Product: cartilage matrix protein #status predicted <MAT>
F:43-210/Domain: von Willebrand factor type A repeat homology <VMA1>
F:231-266/Domain: EGF homology <EGF>
F:277-441/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 11.3%; Score 243.5; DB 2; Length 500;
Best Local Similarity 34.3%; Pred. No. 9.3e-10;
Matches 60; Conservative 29; Mismatches 75; Indels 11; Gaps 2;

QY 31 PRG-----DLMFLDSSASVSHYEFNRVREFVQGLVAPLPLGTGALRASLVHVGSRP 82
DB 34 PRCHLQRTPTDLVFDVDSRSRVPREFEKVFLQVIESLDVGNATRVGLVNTASTV 93
QY 83 YTEFPQHSGEAAQDAVRASQRMGDHTHTGLALVYAKEQLFAEASGAR--PGVPKYV 139
DB 94 KEFPPLRAGSKASLQAVVRIQPLSTGVTGLALPFAIRKALSDAGGARSPDISKTV 153
QY 140 VMTDGGSSDPVPGPMQELKDIGTVFYVSTGRGNFLISAASAPAEKHLHFVD 194
DB 154 IVTVDGRPDSDVSRASAGIELFALGLGRVDRATKLQIASPEQDEHDVVE 208

RESULT 10
A37797

collagen alpha 3(VI) chain precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Aug-2004
C:Accession: A37797; A34270; A32674
R:Doliana, R.; Bonaldo, P.; Colombatti, A.
J. Cell Biol. 111, 2197-2205, 1990
A>Title: Multiple forms of chicken alpha3(VI) collagen chain generated by alternative sp
A:Reference number: A37797; MUID:91035630; PMID:1977751
A:Accession: A37797
A:Molecule type: mRNA
A:Residues: 1-253;312-321;434-453;504-518;635-655;704-717;832-853 <DOL>
A:Cross-references: UNIPROT:P15989; UNIPARC:UPI0000173C31; UNIPARC:UPI0000173C32; UNIPAR
M24282
R:Bonaldo, P.; Russo, V.; Buccitelli, F.; Doliana, R.; Colombatti, A.
Biochemistry 29, 1245-1254, 1990
A>Title: Structural and functional features of the alpha3 chain indicate a bridging role
A:Reference number: A34270; MUID:90212613; PMID:2322559
A:Accession: A34270
A:Molecule type: mRNA
A:Residues: 224-2871 <BON>
A:Cross-references: UNIPARC:UPI0000173C38; GB:M24282
A>Note: the authors translated the codon TTC for residue 1916 as Leu and TTC for residue
J. Biol. Chem. 264, 20235-20239, 1989
A>Title: The carboxyl termini of the chicken alpha3 chain of collagen VI is a unique mo
A:Reference number: A32674; MUID:90062147; PMID:2584214
A:Accession: A32674
A:Molecule type: mRNA
A:Residues: 2151-2199;2792-3137 <BO2>
A:Cross-references: UNIPARC:UPI0000173C39; UNIPARC:UPI0000173C3A; GB:M24282
C:Genetics: C
A:Introns: 30/1; 236/1; 437/1; 638/1; 838/1
C:Superfamily: collagen VI
C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; extracel
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-3137/Product: collagen alpha 3(VI) chain #status predicted <MAT>
F:26-2043/Domain: non-collagenous #status predicted <NMC>
F:36-202/Domain: von Willebrand factor type A repeat homology <VM01>
F:39-404/Domain: von Willebrand factor type A repeat homology <VM02>
F:442-607/Domain: von Willebrand factor type A repeat homology <VM03>
F:642-807/Product: collagen factor type A repeat homology <VM04>
F:840-1004/Domain: von Willebrand factor type A repeat homology <VM05>
F:1033-1197/Domain: von Willebrand factor type A repeat homology <VM06>
F:1237-1400/Domain: von Willebrand factor type A repeat homology <VM07>
F:1439-1604/Domain: von Willebrand factor type A repeat homology <VM08>
F:1639-1804/Domain: von Willebrand factor type A repeat homology <VM09>
F:1838-2010/Domain: von Willebrand factor type A repeat homology <VM10>
F:2043-2378/Domain: collagenous #status predicted <COL>
F:2045-2047/Region: cell attachment (R-G-D) motif
F:2153-2155/Region: cell attachment (R-G-D) motif
F:2153-2161/Region: cell attachment (R-G-D) motif
F:2379-3137/Domain: non-collagenous #status predicted <CNC>
F:2405-2577/Domain: von Willebrand factor type A repeat homology <VM11>
F:2623-2806/Domain: von Willebrand factor type A repeat homology <VM12>
F:2803-2846/Domain: platelet glycoprotein Ib-like #status predicted <GPI>
F:2954-3039/Domain: fibronectin type III repeat homology <FN3>
F:3072-3122/Domain: animal knitin-type proteinase inhibitor homology <BPI>
F:3017,2084,2436,2563,2683,2867,2920,3003/Binding site: carbohydrate (Asn) (covalent

Query Match 11.1%; Score 239; DB 2; Length 3137;
Best Local Similarity 33.5%; Pred. No. 2e-08;
Matches 76; Conservative 31; Mismatches 106; Indels 14; Gaps 6;

QY 2 LPMTA-LGIALSLRLALRSAGERPAPASAPR---GDLMFLDSSASVSHYEFNRVREF 56
DB 7 LPLAAILGLLSS---GFCVGAQ---QAAYNVAVADIIFLVDSWSIGKEHFLVREF 60
QY 57 VQGLVAPLPLGTGALRASLVHVGSRPYTEFPFGQHSGEAAQDAVRASQRMGDHTHTGLA 116
DB 61 LVDVYKALDVGNDRRFALVQPSGNPHTEFQNTYPSNODVLSHTANMPYMGSGSKTGK 120
QY 117 LVYAKQQLFAEASGAR--PGVPKYVLMVTDGSSDPVPGPMQELKDIGTVFYVSTGRGN 174

```

Db      121 LEVLINHLTKAAGSRASEGPGYIVLTLDGGQDDVALPSSVLSAHNMILAVGVQDAV 180
Qy      175 FLELSAAASAPAEKHL-HFVDVDDLHIITVOELRGSILDAMPQQLHA 220
         ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      181 EGELEKEIASRPDTHLFNLENFTALHGIVGLDVASVRTSMTEPOAGA 227

RESULT 11
A37979
cartilage matrix protein precursor - human
C.Species: Homo sapiens (man)
C.Date: 12-Jul-1991 #sequence_revision 12-Jul-1991 #text_change 09-Jul-2004
CAccession: A37979; B37979
R.Jenkins, R.N.; Osborne-Lawrence, S.L.; Sinclair, A.K.; Eddy Jr., R.L.; Byers, M.G.; Sh
J. Biol. Chem. 265, 19624-19631, 1990
A.Title: Structure and chromosomal location of the human gene encoding cartilage matrix
A.Reference number: A37979; PMID:91060568; PMID:2246248
A.Accession: A37979
A.Molecule type: DNA
A.Residues: 1-496 <IEN>
A.Cross-references: UNIPROT:P21941; UNIPARC:UPI000004FLBD; GB:J05667
A.Accession: B37979
A.Molecule type: mRNA
A.Residues: 157-290, 'L', 292-496 <JE2>
A.Cross-references: UNIPARC:UPI000016AEB8; GB:M55683; GB:J05666; GB:J05667; NID:g180651
C.Genetics:
A.Gene: GDB:CRTM
A.Cross-references: GDB:127280; OMIM:115437
A.Map position: lp35-lp35
A.Intron: 32/1, 147/3, 222/1, 264/1, 403/1, 454/1, 481/1
C.Complex: homotrimer
C.Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repeat
C.Keywords: glycoprotein; homotrimer
F.1-22/Domain: signal sequence #status predicted <SIG>
F.23-496/Product: cartilage matrix protein #status predicted <MAT>
F.39-206/Domain: von Willebrand factor type A repeat homology <WMA1>
F.227-262/Domain: EGF homology <EGF>
F.273-437/Domain: von Willebrand factor type A repeat homology <WMA2>
F.76,344/Binding site: carbohydrate (Asn) (covalent) #status predicted
F.221-238,234-247,249-262/Disulfide bonds: #status predicted

Query Match          10.8%; Score 233.5; DB 2; Length 496;
Best Local Similarity 35.0%; Pred. No. 4.7e-09;
Matches    62; Conservative   27; Mismatches    87; Indels     1; Gaps     1;

Qy      34 DIMFLDSSASVSHPFSRVREFVGOLVAAPLPILGALRALSLVHNGSRPYTEPFQGHSS 93
         ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      275 DLVFLIDGSSSVPEPNELVKFTISOVIDTLDVSKLAQVGLVGYSSVRQEFPLGRFT 334

Qy      94 GEAAQADVASAQRMGDTHTGTIALVTAKQLFAEASGARPGVPKYLVWTDDGSSDPVGP 153
         :| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      335 KDIDKAIVRMMSVMEKKTMTGAALKYLINDSFYTVSGSARPGAKXGIYPTDRSQDIYND 394

Qy      154 PMOELDLGVTFIVSTGRCNFLELSAAASP-AEKHLHFVDVDDLHIITVOELRGS 209
         ::||| | | | | | | | | | | | | | | | | | | | | | | |
Db      395 AAKKADLGKMPAVGVGNVDELRERIASPEVAHEFYTADFKTINOIGKKLOKKI 451

RESULT 12
CGHU3A
collagen alpha 3(VI) chain precursor [validated] - human
N.Contains: collagen alpha 3(VI) chain, splice form A9/N10(-)
C.Species: Homo sapiens (man)
C.Date: 21-Nov-1993 #sequence_revision 12-Nov-1999 #text_change 16-Aug-2004
CAccession: A59140; S13679; S24465; A57083; S28776; S00245; C31952; C29848; S26510; S48
R.Chu, W.L.
submitted to GenBank, May 1998
A.Reference number: A59140
A.Accession: A59140
A.Molecule type: mRNA
A.Status: not connected with conceptual translation
A.Residues: 1-3176 <CHU>
A.Cross-references: UNIPROT:P12111; UNIPARC:UPI000012604F; GB:X52022; NID:g3127925; PIDN

```

R.; Chu, M.L.; Zhang, R.Z.; Pan, T.; Stokes, D.; Conway, D.; Kuo, H.J.; Glanville, R.; M
EMBO J. 9, 385-393, 1990
A>Title: Kossac structure of globular domains in the human type VI collagen alpha-3 cha
itors.
A:Reference number: S13679; MUID:90151612; PMID:1689238
A:Accession: S13679
A:Molecule type: mRNA
A:Residues: 1-30,237-313, 'CWM', 318-322, 'AR', 326-1815, 'FD', 1818-1819, 'ID', 1822-3176 <CHS
A:Cross-references: UNIPARC:UPI0000173C14; EMBL:X52022; NID:93127925
A:Accession: S24465
A:Molecule type: protein
A:Residues: 574-585;965-973, 'X', 975-976;1306-1325;1361-1377;1381-1401;1473-1506, 'X', 150
1962, 'X', 1964-1965;2018-2037;2374-2410;2445-2459;2466-2469, 'X', 2471-2474;2504-2508, 'X'
A:Cross-references: UNIPARC:UPI0000173C15; UNIPARC:UPI0000173C16; UNIPARC:UPI0000173C17
C1C; UNIPARC:UPI0000173C1D; UNIPARC:UPI0000173C1E; UNIPARC:UPI0000173C1F; UNIPARC:UPI00
10000173C25; UNIPARC:UPI0000173C26; UNIPARC:UPI0000173C27
R.; Zandi, S.; Doliana, R.; Segat, D.; Bonaldo, P.; Colombatti, A.
J. Biol. Chem. 267, 24082-24089, 1992
A>Title: The human type VI collagen gene, mRNA and protein variants of the alpha3 chain
A:Reference number: S28776; MUID:93054780; PMID:1339440
A:Accession: A57083
A:Molecule type: DNA
A:Residues: 310-328 <ZAN>
A:Cross-references: UNIPARC:UPI0000173C28
A:Accession: S28776
A:Molecule type: mRNA
A:Residues: 32-126, 'AK', 129-136, 'L', 138-236 <ZAN2>
A:Cross-references: UNIPARC:UPI0000173C29; GB:S49432; NID:9260266; PIDN:AA824261.1; PID
R.; Chu, M.L.; Mann, K.; Deutzmann, R.; Pribula-Conway, D.; Hsu-Chen, C.C.; Bernard, M.P.
Eur. J. Biochem. 169, 309-317, 1987
A>Title: Characterization of three constituent chains of collagen type VI by peptide se
A:Reference number: S00126; MUID:88029444; PMID:3365927
A:Accession: S00245
A:Molecule type: mRNA; protein
A:Residues: 2024-2046;2092-2156, 'R', 2203-2208, 'X', 2210-2211, 'X', 2213-2227;2228-2251;231
A:Cross-references: UNIPARC:UPI000016A705; UNIPARC:UPI0000173C29; UNIPARC:UPI0000173C2A
ID:g1335034
A>Note: The mRNA portion of the sequence corresponds to residues 2092-2157
R.; Chu, M.L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutzmann, R.; Timpl, R.
J. Biol. Chem. 263, 18601-18606, 1988
A>Title: Amino acid sequence of the triple-helical domain of human collagen type VI.
A:Reference number: A31952; MUID:89066644; PMID:33198591
A:Accession: C31952
A:Molecule type: mRNA
A:Residues: 2038-2373 <CH4>
A:Cross-references: UNIPARC:UPI0000173C2E; GB:J04211; GB:M20778
A>Note: Parts of this sequence were determined by protein sequencing
R.; Well, D.; Mattei, M.G.; Passage, E.; Van Cong, N.; Pribula-Conway, D.; Mann, K.; Deut
Am. J. Hum. Genet. 42, 435-445, 1988
A>Title: Cloning and chromosomal localization of human genes encoding the three chains
A:Reference number: A29848; MUID:88161046; PMID:3348212
A:Accession: C29848
A:Molecule type: mRNA
A:Residues: 2092-2151 <NEI>
A:Cross-references: UNIPARC:UPI000016A720; GB:M27449; NID:9291919; PIDN:AAA52057.1; PID
A>Note: part of this sequence was determined by protein sequencing
R.; Jander, R.; Rautenberg, U.; Glanville, R.W.
Eur. J. Biochem. 133, 39-46, 1983
A>Title: Further characterization of the three polypeptide chains of bovine and human s
A:Reference number: S26506; MUID:83209648; PMID:6852033
A:Accession: S26510
A:Molecule type: protein
A:Residues: 'SALAGVAGV' <JAN>
A:Cross-references: UNIPARC:UPI0000173C2F
A>Note: this sequence cannot be reliably placed and probably represents the results fro
R.; Mayer, U.; Poeschl, E.; Nischt, R.; Specker, U.; Pan, T.C.; Chu, M.L.; Timpl, R.
Eur. J. Biochem. 225, 573-580, 1994
A>Title: Recombinant expression and properties of the Kunitz-type protease-inhibitor mo
A:Reference number: S48709; MUID:95045506; PMID:755281
A:Accession: S48709
A:Molecule type: mRNA
A:Residues: 'KRAWTFLLCAGRALA', 3102-3176 <MAY>
A:Cross-references: UNIPARC:UPI0000173C30

A>Note: engineered sequence to allow isolation of the Kunitz-type proteinase inhibitor h R.Arnoux, B.; Merigau, K.; Salucljan, P.; Norris, F.; Norris, K.; Bjorn, S.; Olsen, O., submitted to the Brookhaven Protein Data Bank, August 1994

A:Reference number: A52812; PDB:KNT

A:Contents: annotation: X-ray crystallography, 1.6 angstroms, residues 3106-3160

A>Note: engineered sequence expressed in *Saccharomyces cerevisiae* strain mt-663

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C) lines are 5-hydroxylated and subsequently O-glycosylated.

C:Comment: The fibronectin type III repeat homology domain may be released during process

C:Genetics:

A:Gene: GDB:COL6A3

A:Cross-references: GDB:119066; OMIM:120250

A:Map position: 2q37.3-2q37.3

C:Complex: type VI collagen is a heterotrimer of one alpha 1(VI) chain (see PIR:CGH1A), associations among trimer amino- and carboxyl-terminal domains (with disulfide bonds).

C:function:

A:Description: structural component of extracellular tissue microfibrils associated with C:Superfamily: collagen VI

C:Keywords: alternative splicing; blocked amino end; cell binding; coiled coil; extrace

F:1-25/Domain: signal sequence #status predicted <Sig>

F:26-3176/Product: collagen alpha 3(VI) chain #status predicted <MAT1>

F:26-3037/Domain: amino-terminal nonhelical #status predicted <ANH>

F:26-30,237-3176/Product: collagen alpha 3(VI) chain, splice form A9/N10(-) #status pred

F:37-203/Domain: von Willebrand factor type A repeat homology <VM01>

F:240-405/Domain: von Willebrand factor type A repeat homology <VM02>

F:443-608/Domain: von Willebrand factor type A repeat homology <VM03>

F:637-802/Domain: von Willebrand factor type A repeat homology <VM04>

F:835-999/Domain: von Willebrand factor type A repeat homology <VM05>

F:1027-1191/Domain: von Willebrand factor type A repeat homology <VM06>

F:1231-1394/Domain: von Willebrand factor type A repeat homology <VM07>

F:1433-1599/Domain: von Willebrand factor type A repeat homology <VM08>

F:1637-1802/Domain: von Willebrand factor type A repeat homology <VM09>

F:1836-2005/Domain: von Willebrand factor type A repeat homology <VM10>

F:2038-2373/Region: interrupted helical

F:2040-2042/Region: cell attachment (R-G-D) motif

F:2136-2138/Region: cell attachment (R-G-D) motif

F:2148-2150/Region: cell attachment (R-G-D) motif

F:2154-2156/Region: cell attachment (R-G-D) motif

F:2370-2372/Region: cell attachment (R-G-D) motif

F:2370-3176/Domain: carboxyl-terminal nonhelical #status predicted <CNH>

F:2400-2571/Domain: von Willebrand factor type A repeat homology <VM11>

F:2617-2800/Domain: von Willebrand factor type A repeat homology <VM12>

F:2865-2986/Region: alanine/lysine/proline/threonine/valine-rich repeats

F:2987-3072/Domain: fibronectin type III repeat homology <FN3>

F:3111-3161/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F:26/Modified site: pyroglutamate carboxylic acid (Gln) (in mature form) #status predicted

F:108,116,202,251,2079,2331,2558,2677,2861,3036/Binding site: carboxylate (Asn) (coval

F:2100,2206,2239,2316,2319/Modified site: 4-hydroxyproline (Pro) #status experimental

F:2103,2209,2212,2332,2337/Binding site: carboxylate (Lys) (covalent) #status experimen

F:2103,2209,2212,2332,2337/Modified site: 5-hydroxylysine (Lys) #status experimental

F:3111-3161,3120-3144,3136-3157/Disulfide bonds: #status predicted

Query Match 10.5%; Score 225.5; DB 2; Length 3176;

Best Local Similarity 24.8%; Pred. No. 1.8e-07;

Matches 81; Conservative 52; Mismatches 143; Indels 51; Gaps 8;

34 DIMFLDSSASVSHYEFKRVREPVGQVLAFLPLGTALPASLVHVGSRPYTEPFPGQHS 93

39 DILFLVDSKWTIGEEHFLVRELYVAVSLAVGENDFHALVQFNGNHTTELLNTYRT 98

94 GEAADAVPASQRMGDHTGLAVYAKQLFAEASGAR--PGVPKLVVWVTGGSSDPV 151

99 KQEVLSHINMSYIGTNGTGKGLFYIMQSHLTKAAGSRAQGDVPQIVLVLTGHSKDL 158

152 GEPMDLKLQGLVTVFVSGRGNFLELSAASAPAEKHL-HFVDVDDLHIYVELRGSL 210

159 ALPSALAKADVAVFVIGVEDADEGALKETASEPLMMHFNENFTSLDHYGNLVSCH 218

211 DMMRPOOLATETTSQGFRLAMPPLLTADSGYVLEIVSQAQGAARQQLPENATDWIM 270

219 SSVSPERADDTFLKD-----ITAQDSADITFLID----- 248

Query 271 AGLDPEPTYDVAVLPESNRLRPQILRVKTRPEEAGEPRITYSHARPSTL-----RWSG 325

249 -SSNVTGVSVNFVAVLIDFLVNL--EKLPTQOIRVG--VVOFSDEPRMTSLDPTYSIK 302

326 APALGSAALGYHVOFGPLRGGAQRY 352

303 AQVLGAVRLGTF-----AGSELANI 322

RESULT 13

T46488

hypothetical protein DKFZp434J065.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004

C:Accession: T46488

R:UniProt, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S. submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23035

A:Accession: T46488

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-741 <AAA>

A:Cross-references: UNIPROT:Q8ND56; UNIPARC:UPI000016ACGB; EMBL:AL137638

C:Experimental source: adult testis; clone DKFZp434J065

C:Genetics:

A>Note: DKFZp434J065.1

Query Match 10.2%; Score 220; DB 2; Length 741;

Best Local Similarity 31.9%; Pred. No. 7.1e-08;

Matches 59; Conservative 36; Mismatches 84; Indels 6; Gaps 4;

34 DIMFLDSSASVSHYEFKRVREPVGQVLAFLPLGTALPASLVHVGSRPYTEPFPGQHS 93

440 DILFVLDGSKSIGSENFVEVKQFVIGLIDSLTISPKAAVGLQVSTQVHTEFTLRNFS 499

94 GEAADAVPASQRMG-DHTGLAVYAKQLFAEASGAR--GVPKLVVWVTGGSSD 149

500 ADMKKAV-AHKVYKSGSMGLALKMFERSFTGEGARPLSTRPRAIVFTGRRAD 558

150 PVGPPMDLKLQGLVTVFVTSRGNFLELSAASAPAEKHLFV-DVDDLHIYVELRGS 208

559 DVEWASAKKANGITMVAVGKALBEELQETLASEPTNHLFYADFSIMDSISRLKKG 618

209 IUDAM 213

619 ICEAL 623

RESULT 14

A55348

integrin alpha-1 - chicken (fragment)

C:Species: Gallus gallus (chicken)

C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004

C:Accession: A55348

R:Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.B. J. Biol. Chem. 269, 22811-22816, 1994

A:Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1

A:Reference number: A55348; MUID:94357930; PMID:7521332

A:Accession: A55348

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-272 <KER>

A:Cross-references: UNIPROT:Q42094; UNIPARC:UPI000017A14E; GB:U10114

F:55-230/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 9.7%; Score 209; DB 2; Length 272;

Best Local Similarity 33.6%; Pred. No. 1.2e-07;

Matches 73; Conservative 31; Mismatches 85; Indels 28; Gaps 10;

34 DIMFLDSSASVSHYEFKRVREPVGQVLAFLPLGTALPASLVHVGSRPYTEPFPGQHS 93

57 DIVIVLDGNSI--YPMESVTAFNLNLMIDIGFQQTGVIGVIGQTVVHAFYNTYST 114

```
QY 94 GEAADAVASQRMG-DTHITGLALVYAKEQLFAEASGARPGVPKVLVWVTGGSSDPVG 152
DB 115 TEEVMDALRIQRCGTQMTALGIDTAEEAFTENHARGRGVQKVMVITDGEHSD--N 172
QY 153 PMQEL-----KDLGTVFIV-----STGRNPL-----ELSAASAPAEKHLHFVDVD- 197
DB 173 YRLQEVIDKCEDENIQRFAIALIGSYSRNLSTKFEVEIKSIASKPTEK--HFNVSDE 230
QY 198 --LHIIVQELRGSII--LDAMPPOQLHA--TEITSSGF 228
DB 231 LALVTIVEALGERIFALEATTDQAAAFEMWSQAGF 267
```

RESULT 15

```
A45226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45226
R:Brlesewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Cross-references: UNIPROT:P56199; UNIPARC:UPI000012D9EB
A:Experimental source: hepatoblastoma cell line HepG2
A>Note: sequence extracted from NCBI backbone (NCBI:P124326)
F:142-317/Domain: von Willebrand factor type A repeat homology <WMA1>
```

```
Query Match 9.7%; Score 208; DB 2; Length 1151;
Best Local Similarity 27.7%; Pred: 8.9e-07;
Matches 78; Conservative 44; Mismatches 104; Indels 56; Gaps 11;
```

```
QY 34 DLMFLDSSASVSHYEFSSRVREFVQLVAPLIGTALRASLVHSGRPYEPFGQHSS 93
DB 144 DIVIVLDGNSNI--YPMDSVTAFLNDLKRMDIGPKQTQVGIQYGENVTHEFNINKYS 201
QY 94 GEAADAVASQRMG-DTHITGLALVYAKEQLFAEASGARPGVPKVLVWVTGGSSD--P 150
DB 202 TEEVMDALRIQRCGTQMTALGIDTAEEAFTENHARGRGVQKVMVITDGEHSDNR 261
QY 151 VCPMQLDGLGTVFIV-----STGRNPL-----ELSAASAPAEKHLHFVDVD--- 197
DB 262 LKKVIOCCBENIQRFAIALIGSYSRNLSTKFEVEIKSIASKPTEK--HFNVSDELA 319
QY 198 LHIIVQELRGSII--LDAMPPOQLHA--TEITSSGFRLAMPPLLTADSGYVLELVSAOP 253
DB 320 LVTIVKTLGERIFALEATTDQAAAFEMWSQAGF-----SAHY----- 358
QY 254 GAARQQLPQNATDWIMAGIDPDTDYDALVPESNRLLRPQ 295
DB 359 -----SQDWMLGAVGAYDWNQVVMQKASQIITIPR 389
```

Search completed: February 13, 2006, 07:49:18
Job time : 63.2107 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:33:59 ; Search time 329.926 Seconds
(without alignments)
893.868 Million cell updates/sec

Title: US-10-699-035a-6

Perfect score: 2154
Sequence: 1 MPMWALGLALSLRLALARS.....RPRPRVPRAPPTGASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : uniprot_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2130.5	98.9	445	2 Q6PCB0_HUMAN	Q6PCB0 homo sapien
2	1657.5	76.9	415	2 Q8R2Z5_MOUSE	Q8R2Z5 mus musculu
3	1653.5	76.8	415	2 Q923K3_MOUSE	Q923K3 mus musculu
4	1653.5	76.8	415	2 Q8C007_MOUSE	Q8C007 mus musculu
5	1643.5	76.3	415	2 Q642A6_RAT	Q642A6 rattus norv
6	1295.5	60.1	281	2 Q715D7_HUMAN	Q715D7 homo sapien
7	1058.5	49.1	233	2 Q9H6J5_HUMAN	Q9H6J5 homo sapien
8	782.5	36.3	203	2 Q8VDV9_MOUSE	Q8VDV9 mus musculu
9	684	31.8	505	2 Q4SCD1_TETNG	Q4SCD1 tetraodon n
10	437.5	20.3	3124	1 COCA1_CHICK	P13944 gallus gall
11	424	19.7	2884	2 Q5VYK2_HUMAN	Q5VYK2 homo sapien
12	424	19.7	3063	2 Q5VYK1_HUMAN	Q5VYK1 homo sapien
13	421.5	19.6	3063	1 COCA1_HUMAN	Q99715 homo sapien
14	411.5	19.1	1723	2 Q4SD22_TETNG	Q4SD22 tetraodon n
15	407	18.9	3119	1 COCA1_MOUSE	Q60847 mus musculu
16	401	18.6	517	2 Q43853_HUMAN	Q43853 homo sapien
17	392.5	18.2	1259	2 Q4RP12_TETNG	Q4RP12 tetraodon n
18	392.5	18.2	1284	2 Q4VXQ5_HUMAN	Q4VXQ5 homo sapien
19	392.5	18.2	1284	2 Q6P159_HUMAN	Q6P159 homo sapien
20	392.5	18.2	1297	2 Q4VXQ4_HUMAN	Q4VXQ4 homo sapien
21	392.5	18.2	1339	1 K1510_HUMAN	Q9P218 homo sapien
22	392	18.2	637	2 Q81VX1_HUMAN	Q81VX1 homo sapien
23	387	18.0	1117	2 Q4RXN8_TETNG	Q4RXN8 tetraodon n
24	383	17.8	2225	2 Q4SXK3_TETNG	Q4SXK3 tetraodon n
25	382.5	17.8	2944	1 CO7A1_CHICK	Q00238 homo sapien
26	380.5	17.7	1888	1 COEAI_HUMAN	P32018 gallus gall
27	380	17.6	1253	2 Q97566_CANFA	Q97566 canis famli
28	373	17.3	2936	2 Q7YR88_CANFA	Q7YR88 canis famli
29	360	16.7	2944	2 Q63870_MOUSE	Q63870 mus musculu
30	357	16.6	1472	2 Q902A0_CHICK	Q902A0 gallus gall
31	339	15.7	1797	1 COEAI_MOUSE	Q80X19 mus musculu

32	327.5	15.2	1557	2 Q4SH63_TETNG	Q4SH63 tetraodon n
33	324	15.0	1796	1 COEAI_HUMAN	Q05707 homo sapien
34	322.5	15.0	534	2 Q4G0W3_HUMAN	Q4G0W3 homo sapien
35	320.5	14.9	1816	2 Q6NT15_BRARE	Q6NT15 brachydantio
36	292.5	13.6	1060	2 Q4SG43_TETNG	Q4SG43 tetraodon n
37	283.5	13.2	639	1 COCA1_RABIT	Q28902 oryctolagus
38	278	12.9	929	1 COCA1_MOTVI	Q91145 notophthalm
39	266	12.3	624	1 MATN4_MOUSE	Q89029 mus musculu
40	259	12.0	622	1 MATN4_HUMAN	Q95460 homo sapien
41	256	11.9	839	2 Q6P3N7_XENTR	Q6P3N7 xenopus tro
42	254	11.8	647	2 Q4S2X7_TETNG	Q4S2X7 tetraodon n
43	253.5	11.8	451	2 Q8N2G3_HUMAN	Q8N2G3 homo sapien
44	253.5	11.8	956	1 MATN2_HUMAN	Q00339 homo sapien
45	253	11.7	688	2 Q4T0K3_TETNG	Q4T0K3 tetraodon n

ALIGNMENTS

RESULT 1
Q6PCB0_HUMAN
ID Q6PCB0_HUMAN PRELIMINARY; PRT; 445 AA.
AC Q6PCB0
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Von Willebrand factor A domain-related protein, isoform 1 (WARP).
GN Name=WARP; ORFNames=RP4-758J18.11-001;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCB1_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feiglold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.W., Marra W.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX Strausberg R.;
RN Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
[3]
RP NUCLEOTIDE SEQUENCE.
RA Hall R.;
RA Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC059409; AAH59409.1; -, mRNA.
DR EMBL; AL391244; CAI22657.1; -, Genomic DNA.
DR Ensembl; ENSG00000179403; Homo sapiens.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002035; WWF_A.
DR Pfam; PF00041; Fn3; 2.
DR Pfam; PF00092; Wmy; 1.

DR PRINTS: PR00453; VMPADOMAIN.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00327; VMA; 1.
 DR PROSITE; PS50853; FN3; 2.
 DR PROSITE; PS50234; VMA; 1.
 SQ SEQUENCE 445 AA; 46804 MW; D9FDE9C0A4DEBD CRC64;

Query Match 98.9%; Score 2130.5; DB 2; Length 445;
 Best Local Similarity 93.9%; Pred. No. 1.8e-137;
 Matches 418; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

QY 1 MLPTALGLALSLRLALARGAERGPPAPGDLMLDSSASVSHVFSRVREVGQL 60
 DB 1 MLPTALGLALSLRLALARGAERGPPAPGDLMLDSSASVSHVFSRVREVGQL 60
 QY 61 VAPPLGTGALRALSLVHVGSRPYTEPFQSHSGEAAQDAVASAQRMDTHTGLALVYA 120
 DB 61 VAPPLGTGALRALSLVHVGSRPYTEPFQSHSGEAAQDAVASAQRMDTHTGLALVYA 120
 QY 121 KEQLFAEASGAPGVPKVLVWTTDGGSSDPVGPMEQLKDLGVTVFIVSTGRGNFLELSA 180
 DB 121 KEQLFAEASGAPGVPKVLVWTTDGGSSDPVGPMEQLKDLGVTVFIVSTGRGNFLELSA 180
 QY 181 AASAPAEKILHFVDDVDDLIIVQELRGSLIDAMRPQQLHATEITSSGFRLLAMPPLTTADS 240
 DB 181 AASAPAEKILHFVDDVDDLIIVQELRGSLIDAMRPQQLHATEITSSGFRLLAMPPLTTADS 240
 QY 241 GYVYELVPSAQPAGARQOLPGNATDWMAGIDPDTDVALVPSNRLRLPQILRYR 300
 DB 241 GYVYELVPSAQPAGARQOLPGNATDWMAGIDPDTDVALVPSNRLRLPQILRYR 300
 QY 301 TR-----PERAGPRITISHAPRSLRVSAAPALGSA 333
 DB 301 TRPGAGPGASGESGAGAPPTQLAALPAPEAGPERIYISHARPSRLVSWAPALGSA 360
 QY 334 ALGYVQFQPLRGGEAORVEVPAGRNCTTLOGIAPCTAVLVYTAFRSGRSALSAKXC 393
 DB 334 ALGYVQFQPLRGGEAORVEVPAGRNCTTLOGIAPCTAVLVYTAFRSGRSALSAKXC 393
 QY 394 TPDGPRPRRPVPAPTPTGTASREP 418
 DB 394 TPDGPRPRRPVPAPTPTGTASREP 418
 QY 421 TPDGPRPRRPVPAPTPTGTASREP 445
 DB 421 TPDGPRPRRPVPAPTPTGTASREP 445

RESULT 2
 QBR225 MOUSE PRELIMINARY; PRT; 415 AA.
 AC QBR225-
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
 DE Von Willebrand factor A domain-related protein (Mus musculus 11 days
 DE pregnant adult female ovary and uterus cDNA, RIKEN full-length
 DE enriched library, clone:5031410123 product:VON WILLEBRAND FACTOR A-
 DE RELATED PROTEIN homolog).
 GN Name=Vwf;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridea; Muridae; Murinae; Mus.
 CX NCB1_TaxId=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Czech II, and FVB/N;
 RC TISSUE=Colon, and
 RC Mammmary tumor metastasized to lung. Tumor arose spontaneously;
 RX MEDLINE=2238857; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Ditschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,

RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McKernan K.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Czech II;
 RC TISSUE=Mammmary tumor metastasized to lung. Tumor arose spontaneously;
 RG NIH GGC Project;
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning";
 RL Meth. Enzymol. 303:19-44 (1999).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilou S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
 RA Kuenli P., Lewis S., Matsumoto Y., Nakai T., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staib J., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guncinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima Y., Mazzarelli U., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690 (2001).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RA The FANTOM Consortium;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes";
 RL Genome Res. 10:1617-1630 (2000).
 RN [7]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,

RA Komu H, Akiyama J, Nishi K, Kitsuana T, Tashiro H, Itoh M,
 RA Suni N, Ishii Y, Nakamura S, Hazama M, Nishigaki T, Harada A,
 RA Yamamoto R, Matsunoto H, Sakaguchi S, Ikegami T, Kashiwagi K,
 RA Fujiwaka S, Inoue K, Togawa Y, Izawa M, Ohara E, Watahiki M,
 RA Okazaki Y, Ishikawa T, Ozawa K, Tanaka T, Matsura S, Kawai U,
 RA Yoneda Y, Muramatsu M, Inoue Y, Kira A, Hayashizaki Y,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771 (2000).
 RN [8].
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Ovary and uterus;
 RA Adachi J., Atzawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hirakawa T., Hirozane T.,
 RA Hori F., Imocani K., Ishii Y., Itoh M., Kagawa I., Kanakawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Komu H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Saeki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Khahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 RN [9].
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RA Director MGC Project;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC026919; AAH26919.1; -; mRNA.
 DR EMBL; AK077240; BAC36703.1; -; mRNA.
 DR EMBL; BC036166; AAH36166.1; -; mRNA.
 DR HSSP; P18614; IMHP.
 DR Ensembl; ENSMUSG00000042116; Mus musculus.
 DR MGI; MGI:2179729; Wval.
 DR CO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00041; Fn3_2.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VMPADOMAIN.
 DR SMART; SM00660; FN3_2.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS50853; FN3; 2.
 DR PROSITE; PSS0234; VWF_A; 1.
 DR PROSITE; PSS0234; VWF_A; 1.
 SQ SEQUENCE 415 AA; 44709 MW; C3887963B2E334FE CRC64;
 Query March 76.9%; Score 1657.5; DB 2; Length 415;
 Best Local Similarity 77.8%; Pred. No.3.8e-105;
 Matches 325; Conservative 31; Mismatches 95; Indels 3; Gaps 1;
 QY 1 MLPTATGALSLRLALARSAGNERPPASAPRGDLMFLDSSASVSHYEFGRVREFVQQL 60
 Db 1 MLPTATGALSLRLALARSAGNERPPASAPRGDLMFLDSSASVSHYEFGRVREFVQQL 60
 QY 61 VAPPLPGALRAASIVHYGSRPYTFEPFGCHSSSGEADADAVRASQNRNGDHTHTGALVYA 120
 Db 61 VAPPLPGALRAASIVHYGSRPYTFEPFGCHSSSGEADADAVRASQNRNGDHTHTGALVYA 120
 QY 61 VAIMSFGGALRASLIVHGSQPHTEFTEDQYSSGGALIDAIRVAPQRNGDINTGTALAYA 120
 Db 61 VAIMSFGGALRASLIVHGSQPHTEFTEDQYSSGGALIDAIRVAPQRNGDINTGTALAYA 120
 QY 121 KEQFLPAEASGARPGVPTKLVWVTDGSSSDPVPMPQCEKLDIGVYFIYSTGRGNLELSA 180
 Db 121 KEQFLPAEASGARPGVPTKLVWVTDGSSSDPVPMPQCEKLDIGVYFIYSTGRGNLELSA 180
 QY 121 KEQFLPAEASGARPGVPTKLVWVTDGSSSDPVPMPQCEKLDIGVYFIYSTGRGNLELSA 180
 Db 121 KEQFLPAEASGARPGVPTKLVWVTDGSSSDPVPMPQCEKLDIGVYFIYSTGRGNLELSA 180
 QY 181 AASAPAEKHLHPVDVDDLIHTIVQELRGSILDMARPOQLHATEITSSGRLAMPPLITADS 240
 Db 181 AASAPAEKHLHPVDVDDLIHTIVQELRGSILDMARPOQLHATEITSSGRLAMPPLITADS 240
 QY 181 AASAPAEKHLHPVDVDDLIHTIVQELRGSILDMARPOQLHATEITSSGRLAMPPLITADS 240
 Db 181 AASAPAEKHLHPVDVDDLIHTIVQELRGSILDMARPOQLHATEITSSGRLAMPPLITADS 240
 QY 241 GYVYELVPSAQPGARARQOQPGNATDWMGLDPTDGYDVALVESNVRLRLPQILNVR 300
 Db 241 GYVYELVPSAQPGARARQOQPGNATDWMGLDPTDGYDVALVESNVRLRLPQILNVR 300
 QY 241 GYVYELVPSAQPGARARQOQPGNATDWMGLDPTDGYDVALVESNVRLRLPQILNVR 300
 Db 241 GYVYELVPSAQPGARARQOQPGNATDWMGLDPTDGYDVALVESNVRLRLPQILNVR 300
 QY 301 TRPEAGGERIVTSHARPRSLRVSNAPALGSAALGYVDFGPIRLGSAORVEVPAGNC 360
 Db 301 TRPEAGGERIVTSHARPRSLRVSNAPALGSAALGYVDFGPIRLGSAORVEVPAGNC 360

Db	301	TL0E0A6PERIVISHARPRSLRVSMPALGPBDSALGTHV0LQ6G0GSLERVEVPAQNS	360
Oy	361	TTTL0GALPAGVAYVTWTAAPRSRRESALSKACCPDPPRRPRPRVPVPAFPPTGTSAREP	418
Db	361	TTV0GLPCTTLYVTATVTAAPRSGR0RALSKACTASGARTR---AFQSMPEAGPREP	415
RESULT 3			
0923K3_MOUSE			
ID	0923K3_MOUSE	PRELIMINARY;	PRT; 415 AA.
AC	0923K3;		
DT	01-DEC-2001 (TREMBlrel. 19, Created)		
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)		
DT	01-MAR-2004 (TREMBlrel. 26, Last annotation update)		
DE	Von Willebrand factor A-related protein.		
GN	Name=Wval; Synonyms=4932416A11Rik, Wap;		
OS	Mus musculus (Mouse)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Eumarchontoglires; Glires; Rodentia; Sciurognathi;		
OC	Muroidea; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
OX	NCBI_TaxID=10090;		
RP	NUCLEOTIDE SEQUENCE.		
FX	MEDLINE=32057805; Pubmed=12062410; DOI=10.1016/S0014-5793(02)02579-6;		
RA	Fitzgerald J., Ting S.T., Bateman J.F.,		
RT	"WAP: A new member of the von Willebrand factor A-domain superfamily		
RL	of extracellular matrix proteins.";		
DR	FEBS Lett. 517:61-66(2002).		
DR	EMBL; AY030094; AAK8350.1; -; mRNA.		
DR	HSSP; P18614; 1MHP.		
DR	Ensembl; ENSMUSG00000042116; Mus musculus.		
DR	MGI; MGI:2179729; 4932416A11Rik.		
DR	MGI; MGI:2179729; Wval.		
DR	GO; GO:0005615; C:extracellular space; TAS.		
DR	InterPro; IPR003961; FM III.		
DR	InterPro; IPR02035; WMF_A.		
DR	Pfam; PF00041; fn3; 2.		
DR	Pfam; PF00092; WMA; 1.		
DR	PRINTS; PR00453; WMFADOMAIN.		
DR	SMART; SM00060; FN3; 2.		
DR	SMART; SM00327; WMA; 1.		
DR	PROSITE; PS50853; FN3; 2.		
DR	PROSITE; PS50234; WMFA; 1.		
SO	SEQUENCE	415 AA; 44737 MW; C3344AECB3FDE431 CRC64;	
Query Match			
Best Local Similarity	76.8%;	Score 1653.5;	DB 2; Length 415;
Matches	324; Conservative	32; Mismatches	59; Indels 3; Gaps 1;
Oy	1	MLPPTALGLALSLRLALARGAERGPPASAPRGIMFLDLSSASVSHYERSRVEPVQOL	60
Db	1	MLFPTASMSLSRLALARSISIERGSYASDPQGLFLDLSSASVSHYERSRVEPVQOL	60
Oy	61	VAPPLPTGALRSLSIVGSRPYEPFPGCHSSGEAAQDAVRAASQMGDTHTGLAVYA	120
Db	61	VATMSFPGALRSLVIVGSGQPHTEFFFDYSSGQALRDALIRVAPQMGDTNIGLALAYA	120
Oy	121	KEQLFAASGARPEVPVYLVAWTDGSSSDVPVPMQELKDLGTVTFIVSGRGNFLLSA	180
Db	121	KEQLFAAEAGARPEVPVYLVAWTDGSSSDVPVPMQELKDLGTVTFIVSGRGNFLLSA	180
Oy	181	AASAPARKHLHFVVDVLDLHIIYOEIRSGSIIIDAMPPOOLHATETITSSGFRILAMPPLTADS	240
Db	181	AASAPARKHLHFVVDVLDLHIIARLEKRSITDAMQPOULHASEVIVSSGFRILSWPPLTADS	240
Oy	241	GYVYLIVPSAQPGAAARQQLPGNATDVIWAGLDPDTDYVALVPESSVALLRPOQLIVR	300
Db	241	GYVYLIVPSGKLATTRROQLPGNATSWTDTLDDPDTDYEVSLIPESNVHLRPOQHVRV	300
Oy	301	TREEDAPERIVISHARPRSLRVSMAALGSAALGTHV0GGLRGGEADRVVEVPAQRNC	360
Db	301	TL0E0A6PERIVISHARPRSLRVSMAALGSDSLGTHV0GLPQ0GSLERVEVPAQNS	360

OY 361 TTLOGIAPGTAVLVVTTAAFRSGRESALSAKACTPDGPRPRPVPRAPTPGTASREP 418
DB 361 TTVQGLTCTTLYLVTTAAFRSGRORALSAKACTASGAKTR--APQSMRPAAGPREP 415

RESULT 4
O8COO7_MOUSE PRELIMINARY; PRT; 415 AA.

AC O8COO7; 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932416A11 product:VON WILLEBRAND FACTOR A-RELATED PROTEIN homolog.
GN Name=Vwa1; Synonyms=4932416A11R1K;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RT Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085560; PubMed=11217851; DOI=10.1038/35055500;
RT Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner T., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochi H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bash G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima U., Mazzarelli J., Mombarts P.,
RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The RIKEN Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subcloning of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Saeki N., Carninci P.,
RA Kono H., Akiyama Y., Nishi K., Katsuna T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai U.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiroaka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai U., Kojima Y., Kondo S., Kono H., Konda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata N., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ database.
DR HSSP; AK030019; BAC26739.1; -, mRNA.
DR EMBL; AK030019; BAC26739.1; -, mRNA.
DR Ensemble; ENSMUSG0000042116; Mus musculus.
DR MGI; MGI:2179729; 4932416A11R1K.
DR GO; GO:0005615; C:cytoregulation; TAS.
DR GO; GO:0005615; C:cytoregulation; TAS.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00041; FN3_2.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3_2.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00853; FN3_2.
DR PROSITE; PS00234; VWFPA; 1.
SQ SEQUENCE 415 AA; 44681 MW; CE137963B7834FE CRC64;

Query Match 76.8%; Score 1653.5; DB 2; Length 415;
Best Local Similarity 77.5%; Pred. No. 7.1e-105;
Matches 324; Conservative 31; Mismatches 60; Indels 3; Gaps 1;

OY 1 MLPTALGALSLRLALRSGRPPASAPRGDMLFLDSSASVSHYFSPVRFVQGL 60
DB 1 MLPTALGALSLRLALRSGRPPASAPRGDMLFLDSSASVSHYFSPVRFVQGL 60

OY 61 VAPLPLGTGALRASLVHVGSRPYTEPFQGHSSGEAADAVALASQRMGDTHTGLALVYA 120
DB 61 VAPLPLGTGALRASLVHVGSRPYTEPFQGHSSGEAADAVALASQRMGDTHTGLALVYA 120

OY 121 KEQLFAEKASGARPGPKLVWTTDGGSSDPVGPPEQELKDLGVYTFYISTGNGNLEISA 180
DB 121 KEQLFAEKASGARPGPKLVWTTDGGSSDPVGPPEQELKDLGVYTFYISTGNGNLEISA 180

OY 181 KEQLFAEKASGARPGPKLVWTTDGGSSDPVGPPEQELKDLGVYTFYISTGNGNLEISA 180
DB 181 KEQLFAEKASGARPGPKLVWTTDGGSSDPVGPPEQELKDLGVYTFYISTGNGNLEISA 180

OY 241 GYVLELVPSAOPGAARQOLPGNATDWIMAGLDPDTDYVALVBSNVRLPOLRVR 300
DB 241 GYVLELVPSAOPGAARQOLPGNATDWIMAGLDPDTDYVALVBSNVRLPOLRVR 300

OY 301 GYVLELVPSAOPGAARQOLPGNATDWIMAGLDPDTDYVALVBSNVRLPOLRVR 300
DB 301 GYVLELVPSAOPGAARQOLPGNATDWIMAGLDPDTDYVALVBSNVRLPOLRVR 300

OY 361 TTVQGLTCTTLYLVTTAAFRSGRESALSAKACTPDGPRPRPVPRAPTPGTASREP 418
DB 361 TTVQGLTCTTLYLVTTAAFRSGRESALSAKACTPDGPRPRPVPRAPTPGTASREP 418

```
RESULT 5
0642A6 RAT PRELIMINARY; PRT; 415 AA.
ID 0642A6 RAT PRELIMINARY; PRT; 415 AA.
AC 0642A6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Von Willebrand factor A domain-related protein.
GN Name=RGD1311476_predicted;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Kidney.
RC NIH MGC Project;
RX Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
RA EMBL:BC081983; AAH81983.1; -, mRNA.
RA InterPro: IPR003961; FN_III.
RA InterPro: IPR002035; VWF_A.
RA Pfam: PF00041; FN3_2.
RA Pfam: PF00092; VWA_1.
RA PRINTS: PR00453; VWFADOMAIN.
RA SMART: SM00060; FN3_2.
RA SMART: SM00327; VWA_1.
RA PROSITE: PS00853; FN3_2.
RA PROSITE: PS0234; VWF_A_1.
SQ SEQUENCE 415 AA; 44832 MW; B3A0A0268DB06D4E CRC64;

Query Match 76.3%; Score 1643.5; DB 2; Length 415;
Best Local Similarity 77.8%; Pred. No. 3.5e-104;
Matches 325; Conservative 31; Mismatches 59; Indels 3; Gaps 1;
```

```
QY 181 AASAPAEKHLHFVDDVDDLIITVOELRGSLIDAMPPOQLATEITSSGFLAMPPLITADS 240
DB 181 AASAPAEKHLHFVDDVDDLIITVOELRGSLIDAMPPOQLATEITSSGFLAMPPLITADS 240
QY 241 GYVVELEVPASQAGARROQLPGNATDWMAGIDPTDVTVALVPESSNRLRLPQILRYR 300
DB 241 GYVVELEVPASQAGARROQLPGNATDWMAGIDPTDVTVALVPESSNRLRLPQILRYR 300
QY 301 TRPEEAGEPRIVISHAPRSRLSVSWAPALGAAAGYHVFQPLRGGEAQRYVEPAGNRC 360
DB 301 TLQEEAGEPRIVISHAPRSRLSVSWAPALGAAAGYHVFQPLRGGEAQRYVEPAGNRC 360
QY 361 TTLQGLAPCTAVLVVYTAAPFRSGESALSAKACTDGPFRPRRPVPRATPTGTASEP 418
DB 361 TTLQGLAPCTAVLVVYTAAPFRSGESALSAKACTDGPFRPRRPVPRATPTGTASEP 418

RESULT 6
07LSD7 HUMAN PRELIMINARY; PRT; 281 AA.
ID 07LSD7 HUMAN PRELIMINARY; PRT; 281 AA.
AC 07LSD7;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE WARP protein (Fragment).
GN Name=WARP;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Pancreas;
RC NIH MGC Project;
RX Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RA EMBL:BC003543; AAH03543.2; -, mRNA.
RA InterPro: IPR003961; FN_III.
RA Pfam: PF00041; FN3_2.
RA SMART: SM00060; FN3_1.
RA PROSITE: PS00853; FN3_2.
RA NON_TER 1
SQ SEQUENCE 281 AA; 29628 MW; 350CC24590791BD4 CRC64;

Query Match 60.1%; Score 1295.5; DB 2; Length 281;
Best Local Similarity 90.4%; Pred. No. 1.4e-80;
Matches 254; Conservative 0; Mismatches 0; Indels 27; Gaps 1;
```

```
Db      1 VIVITGRGNFLELSAAAPAKHHPVDVDDLIITVDELKSIIDARKPOOLHATEIT 60
Qy      225 SSGFRLAMPPLLTADSGYVLELVPSAQPGAAARQQLPGNATDWIAGIDPTDYVALV 284
Db      61 SSGFRLAMPPLLTADSGYVLELVPSAQPGAAARQQLPGNATDWIAGIDPTDYVALV 120
Qy      285 PSSNVLLRPQILRVTR-----PESAGPERIVISHAR 317
Db      121 PSSNVLLRPQILRVTRPGEAGPGASGPGAGAPATQALPAPEEAGPERIVISHAR 180
Qy      318 PSSLRVSWAPALGSAALGYHVOFGPLRGEAORVVPAGRNCTTQGLAPGAYLVVT 377
Db      181 PSSLRVSWAPALGSAALGYHVOFGPLRGEAORVVPAGRNCTTQGLAPGAYLVVT 240
Qy      378 AAFRSGRESALSAACTPDGPRPRPPVPRAPTGTASREP 418
Db      241 AAFRSGRESALSAACTPDGPRPRPPVPRAPTGTASREP 281
```

RESULT 7

```
Q9H6J5_HUMAN
ID Q9H6J5 HUMAN PRELIMINARY; PRT; 233 AA.
AC Q9H6J5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ22215.
OS Homo sapiens (Human) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
OC Homo
NCBI_TaxID=9606;
(1)
RN NUCLEOTIDE SEQUENCE.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono T., Sugano S.,
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025868; BAB15264.1; -; mRNA.
DR HSSP; Q96KE7; 1FNA.
DR Ensembl; ENSG00000179403; Homo sapiens.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS50853; FN3; 2.
SQ SEQUENCE 233 AA; 24473 MW; B2CC118FCSB7EC50 CRC64;
```

Query Match 49.1%; Score 1058.5; DB 2; Length 233;

Best Local Similarity 88.4%; Pred. No. 1.8e-64; Indels 27; Gaps 1;

Matches 206; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

```
Qy      213 MRPOOLHATEITSSGFRLAMPPLLTADSGYVLELVPSAQPGAAARQQLPGNATDWIAG 272
Db      1 MRPOOLHATEITSSGFRLAMPPLLTADSGYVLELVPSAQPGAAARQQLPGNATDWIAG 60
Qy      273 LDPDIDYVALVPSNVLLRPQILRVTR-----PPE 305
Db      61 LDPDIDYVALVPSNVLLRPQILRVTRPGEAGPGASGPGAGAPATQALPAPEE 120
Qy      306 AAPERIVISHARPRSLRVSWAPALGSAALGYHVOFGPLRGEAORVVPAGRNCTTQGL 365
Db      121 AAPERIVISHARPRSLRVSWAPALGSAALGYHVOFGPLRGEAORVVPAGRNCTTQGL 180
Qy      366 LAPGAYLVTVTAAFRSGRESALSAACTPDGPRPRPPVPRAPTGTASREP 418
Db      181 LAPGAYLVTVTAAFRSGRESALSAACTPDGPRPRPPVPRAPTGTASREP 233
```

RESULT 8
Q8VDV9_MOUSE
ID Q8VDV9_MOUSE PRELIMINARY; PRT; 203 AA.
AC Q8VDV9;

```
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 4932416A1IRik protein.
GN Name=Wval; Synonyms=4932416A1IRik;
OS Mus musculus (Mouse) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
(1)
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altchul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Mullighy S.J.,
RA Raha S.S., Lochellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.W., Sodergren E.U., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.T., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.W., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002) .
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Czech II;
RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020136; AAH02136.1; -; mRNA.
DR Ensembl; ENSMUSG00000042115; Mus musculus.
DR MGI; MGI:2179729; 4932416A1IRik.
DR MGI; MGI:2179729; Wval.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS50853; FN3; 2.
KW Repeat.
SQ SEQUENCE 203 AA; 22068 MW; AAD7FDEF37626898 CRC64;
```

Query Match 36.3%; Score 782.5; DB 2; Length 203;

Best Local Similarity 73.3%; Pred. No. 1.2e-45; Indels 3; Gaps 1;

Matches 151; Conservative 17; Mismatches 35; Indels 3; Gaps 1;

```
Qy      213 MRPOOLHATEITSSGFRLAMPPLLTADSGYVLELVPSAQPGAAARQQLPGNATDWIAG 272
Db      1 MRPOOLHATEITSSGFRLAMPPLLTADSGYVLELVPSAQPGAAARQQLPGNATDWIAG 60
Qy      273 LDPDIDYVALVPSNVLLRPQILRVTRPGEAGPERIVISHARPRSLRVSWAPALGSA 332
Db      61 LDPDIDYVALVPSNVLLRPQILRVTRPGEAGPERIVISHARPRSLRVSWAPALGSA 120
Qy      333 AALGYHVOFGPLRGEAORVVPAGRNCTTQGLAPGAYLVTVTAAFRSGRESALSAA 392
Db      121 SALGYHVOFGPLRGEAORVVPAGRNCTTQGLAPGAYLVTVTAAFRSGRESALSAA 180
Qy      393 CTPDGPRLPRPPVPRAPTGTASREP 418
Db      181 CTAAGATRR---APGSMRPEAGPREP 203
```



```
CC      IsoId=P13944-1; Sequence=Displayed;
CC      Name=Short;
CC      IsoId=P13944-2; Sequence=VSP_001148;
CC      -1- TISSUE SPECIFICITY: Type XII collagen is present in tendons,
CC      ligaments, perichondrium, and perosteum, all dense connective
CC      tissues containing type I collagen.
CC      -1- DOMAIN: This sequence defines five distinct domains, two triple-
CC      helical domains (COL1 and COL2) and three nontriple-helical
CC      domains (NC1, NC2, and NC3).
CC      -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC      each end.
CC      -1- PTM: Prolines at the third position of the tripeptide repeating
CC      unit (G-X-Y) are hydroxylated in some or all of the chains.
CC      -1- PTM: O-glycosylated; glycosaminoglycan or chondroitin-sulfate type
CC      (By similarity).
CC      -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC      interrupted helices (FACIT) family.
CC      -1- SIMILARITY: Contains 18 fibronectin type-III domains.
CC      -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC      -1- SIMILARITY: Contains 4 WFPA domains.
CC      -----
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use as long as its content is in no way modified and this statement is not
CC      removed.
CC      -----
CC      EMBL, D00824; BAA00701.1; -; mRNA.
CC      EMBL, X61024; CAA43358.1; -; mRNA.
CC      EMBL, M17375; AAA48718.1; -; mRNA.
CC      EMBL, J05137; AAA48635.1; -; mRNA.
CC      EMBL, X67327; CAA47744.1; -; mRNA.
CC      FIR, A40020; A40020.
CC      HSP, P56199; 1QC5.
CC      DR      Ensembl; ENSGALG0000015908; Gallus gallus.
CC      DR      InterPro; IPR008160; Collagen.
CC      DR      InterPro; IPR003129; FN_III.
CC      DR      InterPro; IPR002035; VWF_A.
CC      DR      Pfam; PF01391; Collagen; 4.
CC      DR      Pfam; PF00041; fn3; 17.
CC      DR      Pfam; PF00092; VWA; 4.
CC      DR      PRINTS; PR00453; VWFADOMAIN.
CC      DR      SMART; SM00060; FN3; 18.
CC      DR      SMART; SM00210; TSPN; 1.
CC      DR      PROSITE; PS00327; VWA; 4.
CC      DR      PROSITE; PS50853; FN3; 18.
CC      DR      PROSITE; PS50234; WMPA; 4.
CC      KM      Alternative splicing; Cell adhesion; Collagen;
CC      Direct protein sequencing; Extracellular matrix; Glycoprotein;
CC      Hydroxylation; Repeat; Signal; Structural protein.
CC      SIGNAL      1      24
CC      FT      CHAIN      25      3124      Collagen alpha 1(XII) chain.
CC      FT      DOMAIN      25      112      Fibronectin type-III 1.
CC      FT      DOMAIN      139      311      VWFPA 1.
CC      FT      DOMAIN      332      421      Fibronectin type-III 2.
CC      FT      DOMAIN      439      615      VWFPA 2.
CC      FT      DOMAIN      630      718      Fibronectin type-III 3.
CC      FT      DOMAIN      721      809      Fibronectin type-III 4.
CC      FT      DOMAIN      812      902      Fibronectin type-III 5.
CC      FT      DOMAIN      905      993      Fibronectin type-III 6.
CC      FT      DOMAIN      995      1083      Fibronectin type-III 7.
CC      FT      DOMAIN      1086      1175      Fibronectin type-III 8.
CC      FT      DOMAIN      1199      1371      VWFPA 3.
CC      FT      DOMAIN      1386      1472      Fibronectin type-III 9.
CC      FT      DOMAIN      1474      1564      Fibronectin type-III 10.
CC      FT      DOMAIN      1566      1654      Fibronectin type-III 11.
CC      FT      DOMAIN      1655      1745      Fibronectin type-III 12.
CC      FT      DOMAIN      1756      1845      Fibronectin type-III 13.
CC      FT      DOMAIN      1847      1935      Fibronectin type-III 14.
CC      FT      DOMAIN      1937      2026      Fibronectin type-III 15.
CC      FT      DOMAIN      2038      2117      Fibronectin type-III 16.
CC      FT      DOMAIN      2119      2206      Fibronectin type-III 17.
```

```
FT      DOMAIN      2210      2284      Fibronectin type-III 18.
FT      DOMAIN      2327      2500      WMPA 4.
FT      DOMAIN      2524      2716      TSP N-terminal.
FT      REGION      2455      2750      Nonhelical region (NC3).
FT      REGION      2751      2902      Triple-helical region (COL2) with 1
FT      REGION      2903      2945      imperfection.
FT      REGION      2946      3048      Nonhelical region (NC2).
FT      REGION      2946      3048      Triple-helical region (COL1) with 2
FT      REGION      3049      3124      imperfections.
FT      MOTIF      2899      2901      Nonhelical region (NC1).
FT      COMBINS      3086      3096      Cell attachment site (Potential).
FT      COMBINS      3111      3123      Asp/Glu-rich (acidic).
FT      CARBOHYD      32      32      Arg/Lys-rich (basic).
FT      CARBOHYD      797      797      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD      797      797      O-linked (Xyl...) (chondroitin sulfate)
FT      CARBOHYD      890      890      (Potential).
FT      CARBOHYD      890      890      O-linked (Xyl...) (chondroitin sulfate)
FT      CARBOHYD      981      981      (Potential).
FT      CARBOHYD      981      981      O-linked (Xyl...) (chondroitin sulfate)
FT      CARBOHYD      1006      1006      (Potential).
FT      CARBOHYD      1032      1032      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD      1044      1044      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD      1512      1512      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD      1767      1767      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD      2210      2210      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD      2273      2273      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD      2532      2532      N-linked (GlcNAc...) (Potential).
FT      CARBOHYD      2683      2683      N-linked (GlcNAc...) (Potential).
FT      VARREPLIC      25      1188      Missing (in isoform Short).
FT      CONFLICT      1258      1258      /FTId=VSP_001148.
FT      CONFLICT      1264      1264      T -> S (in Ref. 4).
FT      CONFLICT      2759      2759      D -> E (in Ref. 4).
FT      CONFLICT      2803      2803      P -> A (in Ref. 2).
FT      CONFLICT      2977      2977      L -> F (in Ref. 2).
FT      CONFLICT      3075      3076      V -> F (in Ref. 2).
FT      CONFLICT      3124      3124      QP -> AG (in Ref. 3).
SQ      SEQUENCE      3124 AA; 340582 MW; 094285AE7F346CF CRC64;

Query Match      20.3%; Score 437.5; DB 1; Length 3124;
Best Local Similarity 31.1%; Pred. No. 1.2e-20;
Matches 123; Conservative 66; Mismatches 168; Indels 39; Gaps 8;

QY      32      RQDLMFLLDSASVSHTFSPKRVFVGVLPGLGTALPASLVHVSRRPYTEPPQCH 91
DB      437      KADVFLVDGSSYSIGIANFVKRAFLVFLVKSFEISPKRVQISIVQYSDRDPMETSLRKY 496
QY      92      SSGEADQDAVRAASQRMGDTHGLAVYAKGOLFPAASGARPGVYVWTTDGGSDPV 151
DB      497      NKVKTIQALINFPYRGGSTVTKAMTVREKVFVTSKGRPNVPRWMLITDGSDF 556
QY      152      GPPMOLKDLGVTVIVSTGRGNFLESAASAPAEKHLHFV-DVDDHLIIQELRGS1- 209
DB      557      KEPALKADADVEIFAVGVKDAVRLELAISPRFETHVYVEDBDAQRSFELTQSV 616
QY      617      LRIEQLAIRKKSYPVPAKNWFSVDTSDSFVSAAGSEKSYLIRKYV--AIGDEF 674
DB      675      IVSVASSTSVLTNLLFETTYAVSVAIE-----YEDGDGPPLDGETTLEVGA 724
QY      210      -----LDAMR-----DQQLHATEITSSGRLAMPPLITLADSGYVLELVPSAQGAAR 257
DB      617      LRIEQLAIRKKSYPVPAKNWFSVDTSDSFVSAAGSEKSYLIRKYV--AIGDEF 674
QY      258      RQQLPGNMTDMIMAGLDPDTDYDALVPESNVRLLRPQLKVRTRPEEAGGERIVISAR 317
DB      675      IVSVASSTSVLTNLLFETTYAVSVAIE-----YEDGDGPPLDGETTLEVGA 724
QY      318      PSRLR-----VSWAPALGSAAALGYHVFGLRGEAORVYEPAGRNCTTLOGIA 367
DB      725      PRNLRIETBETDTSFVIGTTPAGN--VLRYLVRVPLTGGERRQVTVANSRSTTLRL 782
QY      368      PGTAVLVTVTAAFRSGRESALSAKCTPDGPRPRR 403
DB      783      PDRYEVSVIAEVQSGPGNALNGVAKT-DEVKGNR 817

RESULT 11
```

```

Q5VYK2_HUMAN
ID Q5VYK2_HUMAN PRELIMINARY; PRT; 2884 AA.
AC Q5VYK2;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Collagen, type XII, alpha 1.
GN Name=COL12A1; ORFNames=RP1-238D15.1-003;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OC NCBI_TaxId=9606;
RN NUCLEOTIDE SEQUENCE.
RA Tromans A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RP NUCLEOTIDE SEQUENCE.
RA Smith M.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN NUCLEOTIDE SEQUENCE.
RA Corby N.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL354664; CAH71309.1; -; Genomic DNA.
DR EMBL; AL080250; CAI19897.1; -; Genomic DNA.
DR EMBL; AL096771; CAI19907.1; -; Genomic DNA.
DR EMBL; AL080250; CAH71309.1; JOINED; Genomic DNA.
DR EMBL; AL096771; CAH71309.1; JOINED; Genomic DNA.
DR EMBL; AL354664; CAI19897.1; JOINED; Genomic DNA.
DR EMBL; AL080250; CAI19907.1; JOINED; Genomic DNA.
DR EMBL; AL354664; CAI19907.1; JOINED; Genomic DNA.
DR EMBL; AL096771; CAI19897.1; JOINED; Genomic DNA.
DR EMBL; ENSG0000011799; Homo sapiens.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0005578; Extracellular matrix (sensu Metazoa); IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01391; Collagen_2.
DR Pfam; PF00041; fn3; 18.
DR Pfam; PF00092; VWA; 4.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 18.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; VWA; 4.
DR PROSITE; PS50853; FN3; 18.
DR PROSITE; PS50234; VWFA; 4.
DR Collagen; Extracellular matrix; Repeat; Structural protein.
SQ SEQUENCE 2884 AA; 315869 MW; 2D598F13656454D CRC64;

Query Match 19.7%; Score 424; DB 2; Length 2884;
Best Local Similarity 29.9%; Pred. No. 9.4e-20;
Matches 120; Conservative 68; Mismatches 186; Indels 28; Gaps 8;

QY 32 RQDLMLFLDSSASVSHYEFSSRYAEFGQVLAPLGLTGALRSALVHVGSRPYTFEPFGQH 91
DB 438 KADIVFLVVGSSYSIGIANFVKVAFLVLEVLVKSFEISPNRVQISLVQYSRDPHTFELKKF 497
QY 92 SSGEAAQDAVRSASQMGDTHTGLALVYAKEQLFAASGARPVPVKLVVWTDGSSDPV 151
DB 498 TKVEDIIEAINTFPYVGSGSTNTGKAMTYREKIFVPSKSRSNVPMKMLITDGKSDAF 557
QY 152 GPMOELKGLGVTVFVSTGRGNFLSEASAASAPAEKHILFV-DVNDLHIIVQELGSI- 209
DB 558 RDPATKRSNDVEIFAVGVKDAVRSLEALASPAETHVFTVDPDAFQRISELVQSIQ 617
QY 210 -----LDAMR-----PQQLHATEITSSGFGRLAMPPLLADSGYVLELVPSAGPGAAR 257

```

```

DB 618 LRIEQLAIKKKAVPPDDLSFSSVTSYGFRTNNSPGENVSYHI-----TYKEAG 671
QY 258 RQOL-----PGNATDWMAGLDPDPTDYDALVAVESNVRLLRPLQLRTRPEAGEERIYI 313
DB 672 DDEVTVVERPASTSVLSSLKRETIYLVNVAIEYDGSIFLAGETTHEVGVARNLKV 731
QY 314 SHARPRLSVSWAPALGSAALGVHYVFGPLRGSGAORVVEPAGNCTTLOGLAFGTAYL 373
DB 722 TDETFDSFKITWTQAPG--RVLRVRIIVRPVAGSGESREVTTPPNORRTLEMLIPDTKYE 789
QY 374 VVTVAAPRSGRSALSACACTPDGRRPRRPVPRAPRTGTAS 415
DB 790 VSVIPEYFSGPGTPLTGNAATEE-VRGNPRDL-RVSDPTTST 829

```

RESULT 12

Q5VYK1_HUMAN

ID Q5VYK1_HUMAN PRELIMINARY; PRT; 3063 AA.

AC Q5VYK1;

DT 01-FEB-2005 (TrEMBLrel. 29, Created)

DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Collagen, type XII, alpha 1

GN Name=COL12A1; ORFNames=RP1-238D15.1-001;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;

OC Homo.

OC NCBI_TaxId=9606;

RN NUCLEOTIDE SEQUENCE.

RA Tromans A.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

RP NUCLEOTIDE SEQUENCE.

RA Smith M.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

RN NUCLEOTIDE SEQUENCE.

RA Corby N.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL354664; CAH71310.1; -; Genomic DNA.

DR EMBL; AL080250; CAI19898.1; -; Genomic DNA.

DR EMBL; AL096771; CAI19908.1; -; Genomic DNA.

DR EMBL; AL080250; CAH71310.1; JOINED; Genomic DNA.

DR EMBL; AL354664; CAI19898.1; JOINED; Genomic DNA.

DR EMBL; AL354664; CAI19908.1; JOINED; Genomic DNA.

DR EMBL; AL096771; CAI19898.1; JOINED; Genomic DNA.

DR EMBL; ENSG0000011799; Homo sapiens.

DR GO; GO:0005737; Cytoplasm; IEA.

DR GO; GO:0005578; Extracellular matrix (sensu Metazoa); IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0007155; P:cell adhesion; IEA.

DR GO; GO:006817; P:phosphate transport; IEA.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR003961; FN_III.

DR InterPro; IPR003129; Laminin_G_TSP_N.

DR InterPro; IPR002035; VWF_A.

DR Pfam; PF01391; Collagen_2.

DR Pfam; PF00041; fn3; 18.

DR Pfam; PF00092; VWA; 4.

DR PRINTS; PR00453; VWFADOMAIN.

DR SMART; SM00060; FN3; 18.

DR SMART; SM00210; TSPN; 1.

DR SMART; SM00327; VWA; 4.

DR PROSITE; PS50853; FN3; 18.

DR PROSITE; PS50234; VWFA; 4.

DR Collagen; Extracellular matrix; Repeat; Structural protein.

SQ SEQUENCE 3063 AA; 333147 MW; EA38CAFECEB893D2 CRC64;

```
Query Match      19.7%; Score 424; DB 2; Length 3063;
Best Local Similarity 29.9%; Pred. No. 1e-19;
Matches 120; Conservative 68; Mismatches 186; Indels 28; Gaps 8;

Oy 32 RGDLMFLDSSASVSHYEFRRVEFGVAPLPICGTALRASLVHVSRRPTEPPFGH 91
Db 438 KADIVFLVDGSIIGIANFVKRAFLVLYKSFELSPNRVQSLVQYSRDPTEETLKKF 497
Oy 92 SSGEAAQDAVRAAQRMGDPHTLTALVYAEQLFAEASGAPRGVKKLVMTDGGSDPV 151
Db 498 TKVEDIIEAINTFPYRGSGTNGKAMTYREKIFVPSKSGSRGNVKKMILLIDGSSDAF 557
Oy 152 GPPMELDLQGTIVTIVTSGRGNFELASAAAPAEKHLHFV-DVDDHLIIYQELRGSI- 209
Db 558 RDPALIKNSDVEIIPAVGVKQAVRSELEAIASPPAEITHVFTVEDDAFORISFELTQSIQ 617
Oy 210 -----LDPMR-----POOLATEITSSGFFLAMPPLLTADSGYVLELVSAQPGAR 257
Db 618 LRIBELAIKKKAVVPPDLSFSEVTSYGFKTMSPAGEVNFSTHI-----TYKEANG 671
Oy 258 RQQL-----PGNATDWIMAGLDPDIDYVALVPSNVRLRPOOLRVRTRPEAGEPRIV 313
Db 672 DDEVTVPEPASTSVLSSLKPEETLYLVNVTAEYEDGFSIPLAGEETTEVVGADRNLIKV 731
Oy 314 SHARRSRVSWAPALGSAALGYVQFGLRGGAQRVEVAGNCTTLOGLAGTAYL 373
Db 732 TDETTDSKRTWTQAPG--RVLRYRIYKRPVAGSGSRVETTPPNORRRLTLEMLIDTYKE 789
Oy 374 VTVTAAPRSGRSALSAKACTPDGPRPRPVPRAPTPTGAS 415
Db 790 VSVIPEYPSGPGCTPLTGNAPATE-VKGNPRDL-RVSDPTTST 829

RESULT 13
COCAL HUMAN STANDARD; PRT; 3063 AA.
ID COCAL HUMAN STANDARD; PRT; 3063 AA.
AC Q99715; Q99715; (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN Name=COL12A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCB1_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE (ISOFORMS LONG AND SHORT), AND PARTIAL PROTEIN
RP MEDLINE=97288521; PubMed=9143499; DOI=10.1006/geno.1997.4638;
RA Gerecke D.R., Olson P.F., Koch M., Knoll J.H.M., Taylor R.,
RA Hudson D.L., Champilaud M.-F., Olsen B.R., Burgeson R.E.;
RT "Complete primary structure of two splice variants of collagen XII,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(IX)
RT collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human
RT chromosome 6q12-q13."
RL Genomics 41:236-242(1997).
CC -1- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the perifibrillar matrix (By similarity).
CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment:the final tissue form of collagen XII may contain
CC homotrimers of either isoform Long or isoform Short or any
CC combination of isoform Long and isoform Short;
CC Name=Long;
CC Name=Short;
CC Name=Q99715-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q99715-2; Sequence=VSP_001149;
```

```
CC -1- TISSUE SPECIFICITY: Found in collagen I-containing tissues: both
CC isoform Short and isoform Long appear in amnion, chorion, skeletal
CC muscle, small intestine, and in cell culture of dermal
CC fibroblasts, keratinocytes and endothelial cells. Only the short
CC isoform is found in lung, placenta, kidney and a squamous cell
CC carcinoma cell line.
CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC each end (By similarity).
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains (By
CC similarity).
CC -1- PTM: O-glycosylation of isoform Long; glycosaminoglycan of
CC chondroitin-sulfate type (By similarity).
CC -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC interrupted helices (FACIT) family.
CC -1- SIMILARITY: Contains 18 fibronectin type-III domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -1- SIMILARITY: Contains 4 VWFA domains.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL: U73778; AAC51244.1; -; mRNA.
CC EMBL: U73779; AAD40483.1; -; mRNA.
CC HSSP: P18614; IMHP.
CC DR Ensembl: ENSG00000111799; Homo sapiens.
CC HGNC: HGNC:2188; COL12A1.
CC MIM: 120320; -.
CC GO: GO:0005595; C:collagen type XII, TAS.
CC GO: GO:0001501; P:skeletal development; TAS.
CC InterPro: IPR008160; Collagen.
CC InterPro: IPR003961; FN III.
CC InterPro: IPR002035; VWF_A.
CC Pfam: PF01391; Collagen_4.
CC Pfam: PF00041; Fn3; 18.
CC Pfam: PF00092; VWFA; 4.
CC PRINTS: PR00453; VWFADOMAIN.
CC PROSITE: PS50853; FN3; 18.
CC PROSITE: PS50234; VWFA; 4.
CC KW Alternative splicing; Cell adhesion; Collagen;
CC Direct protein sequencing; Extracellular matrix; Glycoprotein;
CC Hydroxylation; Repeat; Signal; Structural protein.
CC SIGNAL 1 24
CC CHAIN 25 3063 Collagen alpha 1(XII) chain.
CC DOMAIN 25 112 Fibronectin type-III 1.
CC DOMAIN 140 316 VWFA 1.
CC DOMAIN 333 422 Fibronectin type-III 2.
CC DOMAIN 440 616 VWFA 2.
CC DOMAIN 631 719 Fibronectin type-III 3.
CC DOMAIN 722 810 Fibronectin type-III 4.
CC DOMAIN 813 901 Fibronectin type-III 5.
CC DOMAIN 904 993 Fibronectin type-III 6.
CC DOMAIN 995 1083 Fibronectin type-III 7.
CC DOMAIN 1086 1175 Fibronectin type-III 8.
CC DOMAIN 1199 1371 VWFA 3.
CC DOMAIN 1384 1472 Fibronectin type-III 9.
CC DOMAIN 1474 1563 Fibronectin type-III 10.
CC DOMAIN 1565 1652 Fibronectin type-III 11.
CC DOMAIN 1654 1743 Fibronectin type-III 12.
CC DOMAIN 1752 1841 Fibronectin type-III 13.
CC DOMAIN 1843 1931 Fibronectin type-III 14.
CC DOMAIN 1933 2022 Fibronectin type-III 15.
CC DOMAIN 2024 2113 Fibronectin type-III 16.
CC DOMAIN 2115 2202 Fibronectin type-III 17.
CC DOMAIN 2206 2290 Fibronectin type-III 18.
CC DOMAIN 2292 2383 Fibronectin type-III 19.
CC DOMAIN 2385 2474 Fibronectin type-III 20.
CC REGION 2451 2746 TSP N-terminal.
CC REGION 2747 2898 Triple-helical region (COL2) with 1
CC imperfection.
```

FT	REGION	2899	2941	Nonhelical region (NC2).
FT	REGION	2942	3044	Triple-helical region (COL1) with 2 imperfections.
FT	REGION	3045	3063	Nonhelical region (NC1).
FT	MOTIF	862	864	Cell attachment site (Potential).
FT	MOTIF	2781	2781	Cell attachment site (Potential).
FT	MOTIF	2885	2887	Cell attachment site (Potential).
FT	MOD_RES	2944	2944	Hydroxyproline (By similarity).
FT	MOD_RES	2947	2947	Hydroxyproline (By similarity).
FT	MOD_RES	2950	2950	Hydroxyproline (By similarity).
FT	MOD_RES	2959	2959	Hydroxyproline (By similarity).
FT	MOD_RES	2965	2965	Hydroxyproline (By similarity).
FT	MOD_RES	2968	2968	Hydroxyproline (By similarity).
FT	MOD_RES	2971	2971	Hydroxyproline (By similarity).
FT	MOD_RES	2983	2983	Hydroxyproline (By similarity).
FT	MOD_RES	3000	3000	Hydroxyproline (By similarity).
FT	MOD_RES	3003	3003	Hydroxyproline (By similarity).
FT	MOD_RES	3014	3014	Hydroxyproline (By similarity).
FT	MOD_RES	3023	3023	Hydroxyproline (By similarity).
FT	MOD_RES	3026	3026	Hydroxyproline (By similarity).
FT	MOD_RES	3029	3029	Hydroxyproline (By similarity).
FT	CARBOHYD	700	700	N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	798	798	O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT	CARBOHYD	889	889	O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT	CARBOHYD	981	981	O-linked (Xyl...) (chondroitin sulfate) (Potential).
FT	CARBOHYD	1763	1763	N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	2206	2206	N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	2528	2528	N-linked (GlcNAc...) (Potential).
FT	CARBOHYD	2679	2679	N-linked (GlcNAc...) (Potential).
FT	VARSPPLIC	25	1188	Missing (in isoform Short).
SQ	SEQUENCE	3063 AA;	333194 MW;	75FEA78FA8E8293 CRC64;
Query Match		19.6%;	Score 421.5;	DB 1; Length 3063;
Best Local Similarity		27.5%;	Pred. No. 1.5e-19;	
Matches 128;		Conservative 59;	Mismatches 174;	Indels 105; Gaps 9;
QY	32	RDLMFLDSSASVSHYFSRVRFEVGOVLAPLPLGTGLARSLVHVGSRPYTEPFQGH	91	
DB	438	KADIVFLVDGYSIGIANFVKARAFLEVLVKSFEISPNVQISLVYSDPHTEFLTKF	497	
QY	92	SSGEAQAQAVRASAOQMDTHTGALVYAKQQLFAEASGARQGVPLVWMTDGGSSDPV	151	
DB	498	TKVEDIIEIAINTFPYRGGSTNTGKATYREKI FVPSKSRNVKPMILLIDGKSSDAF	557	
QY	152	GPOMELKDLGVTIVSTGRGNFLELSAASAPAEKHLAFV-DVDDLHIYQELRGSI-	209	
DB	558	RDPALIKRNSDVEIFAVGVKDAVRSLELAISPPAETHVFTVBDPAQRISBELTQSI	617	
QY	210	-----LDAMR-----POOLHATEITSSGFRLAMPPL-----LTADSGYV	244	
DB	618	LRIBQELAIKKKAVVPPKDLSEFSEVTSYGFYTNMSPAGEVNSYHITYKEAGDEEVV	677	
QY	245	LE-----LVPSAQP-----GAARROQLPQNAD	267	
DB	678	VBPASTSVVLSLKEPTELYLVNVTAAEYDGSIPLAGETTBVYKAGARNILKVNDEDTD	737	
QY	268	-----WMA-----GLDPDPTYDVALVPSN	288	
DB	738	SFKITMTQAPGRVLCRIIYRPVAGSESEVTPPNQRRRTLENLIPDKRYEVSIVPEVF	797	
QY	289	VLLRPQILRVTRPREAGPERIVISHAPRSIRVSMAPALSGAALGYHVOGFLRGGE	348	
DB	798	SGPGTPTGNMAATEEVGRNPRDLRVSDPTSTKLSWGAFGVVKQ--YLVVYTTVPAQGE	855	
QY	349	AORVEVPAGRNCTTQGLAPGTAIVTVTAAPRSGESALSAKACT	394	
DB	856	TOEVTVRGDTTNTVLOGLKEGTOYALSVTALYASGADALFEGGTT	901	

RESULT 14	Q4SD22_TESTNG	PRT;	1723 AA.
AC	Q4SD22;		
DT	13-SEP-2005 (TREMBlrel. 31, Created)		
DT	13-SEP-2005 (TREMBlrel. 31, last sequence update)		
DT	13-SEP-2005 (TREMBlrel. 31, last annotation update)		
DE	Chromosome 14 SCAF14645, whole genome shotgun sequence.		
GN	ORFNames=GSTENG00020231001;		
OS	Tetradodon nigroviridis (Green puffer).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetradontiformes;		
OC	Tetradontidae; Tetradontidae; Tetradontidae;		
OX	NCBI_TaxID=99883;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Jailon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,		
RA	Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,		
RA	Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,		
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,		
RA	Antouard V., Jubin C., Catroilo L., Pouliou J., De Bernardis V.,		
RA	Bienmont C., Skallit Z., Coutanceau J.P., Guzy J.,		
RA	Cruaud C., Duprat S., Broctier P., Coucanceau J.P., Guzy J.,		
RA	Parra G., Lartier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,		
RA	Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,		
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,		
RA	Laudet V., Schachter V., Queller F., Saurin W., Scarpelli C.,		
RA	Winkler P., Lander E.S., Weissbach J., Roset Crolius H.,		
RT	"Genome duplication in the teleost fish Tetradodon nigroviridis reveals		
RL	the early vertebrate proto-karyotype."		
RL	Nature 431:946-957(2004).		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RG	Genoscope, Whitehead Institute Centre for Genome Research,		
RL	Submitted (FBS-2004) to the EMBL/Genbank/DBS databases.		
CC	-1- CAUTION: The sequence shown here is derived from an		
CC	EMBL/Genbank/DBS whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
CC	EMBL; CA001014645; CAG01460.1; -, Genomic_DNA.		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR003962; FNIII_subd.		
DR	InterPro; IPR02035; WVF_A.		
DR	Pfam; PF00041; fn3; 11.		
DR	Pfam; PF00092; WMA; 3.		
DR	PRINTS; PR00014; ENTYPBIII.		
DR	PRINTS; PR00453; WVFADOMAIN.		
DR	SMART; SMO0060; FN3; 11.		
DR	SMART; SMO0327; WMA; 3.		
DR	PROSITE; PS00853; FN3; 12.		
DR	PROSITE; PS00334; WMA; 3.		
KW	Collagen; Extracellular matrix; Structural protein.		
FT	NON TER		
FT	NON TER		
SQ	SEQUENCE	1723 AA;	187496 MW;
Query Match		19.1%;	Score 411.5;
Best Local Similarity		26.5%;	Pred. No. 3.6e-19;
Matches 128;		Conservative 54;	Mismatches 180;
			Indels 121; Gaps 8;
QY	32	RDLMFLDSSASVSHYFSRVRFEVGOVLAPLPLGTGLARSLVHVGSRPYTEPFQGH	91
DB	436	KADIVFLVDGYSIGIANFVKARAFLEVLVKSFEISPNVQISLVYSDPHTEFLTDSH	495
QY	92	SSGEAQAQAVRASAOQMDTHTGALVYAKQQLFAEASGARQGVPLVWMTDGGSSDPV	151
DB	496	HNLEAVLVALTRFPYRGGSTNTGRAMTVRETFQASRGARAHVPRVITLLIDGKSSDAF	555
QY	152	GPOMELKDLGVTIVSTGRGNFLELSAASAPAEKHLAFV-DVDDLHIYQELRGSI-	189
DB	556	QBPANLRNSDVEIFAVGVKDAVRSLELAISPPAETHVFTVBDPAQRISTELTQSI	615

QY 190 -----LHFVY-----DDLHI 201
 Db 616 LRIEGLQIINGRLVQPRDILFSDVGPFRFASWEINANNVSYLVQFRPTEGDSHYV 675
 QY 202 VOELGSLIUDAMP-----COLNATEIT 224
 Db 676 SMSVSGDVLTALLPHILPTLTREVSVAQYANGTSLPTVGYCTTAEERGSVONLVTEES 735
 QY 225 SSGFELAMPPLLTADSGYVLELVPSAQPGARRQOL-----PGNATDWIMAGLDPDPTDYD 280
 Db 736 POSFVSWRAAFAGAVARRL-----TYOPAGAGAEQLAEFTAGPELITWVLODLPRTTYR 790
 QY 281 VALVESNRVRLRPQLLRVTRPEEAGPERIVTSHARPRLSVASAPALGSAALGTHVQ 340
 Db 791 VTVTEYKGGPGGPGQDTGDTTTEAKSGPGLRVFDQTVSSMKVSWEPAPGD--VLQYRLA 848
 QY 341 FGPIRGEAORVEVAGRNCTTLOGLPATVLTVTAFRSGRSALSAKCTPD--GP 398
 Db 849 YRASGSGRKEVSVGNTAVLKQLQETEDITVSARVNSGLGALKRGATTEEVGP 908
 QY 399 RPR 401
 Db 909 -PR 910
 RESULT 15
 COCAL MOUSE STANDARD; PRT; 3119 AA.
 ID COCAL MOUSE
 AC Q60847; P70322;
 DT 15-JUN-1998 (Rel. 36, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor.
 GN Name=Coll12a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING (ISOFORMS XIIA-1 AND XIIA-2)
 RC STRAIN=C57BL/6J, and Swiss Webster; TISSUE=Skin;
 RX MEDLINE=96170761; PubMed=8601036;
 RA Boehme K., Li Y., Oh P.S., Olsen B.R.;
 RT "Primary structure of the long and short splice variants of mouse collagen XII and their tissue-specific expression during embryonic development."
 RT Dev. Dyn. 204:432-445 (1995).
 RL [2]
 RN PARTIAL NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING (ISOFORMS XIIA-2 AND XIIA-2)
 RP STRAIN=C57BL/6J; TISSUE=Skin fibroblast;
 RC MEDLINE=99348349; PubMed=10419532; DOI=10.1074/jbc.274.31.22053;
 RA Kanda A.W., Reichenberger E., Baur S.T., Karimbox N.Y., Taylor R.W., Olsen B.R., Nishimura I.;
 RT "Structural variation of type XII collagen at its carboxyl-terminal NCI domain generated by tissue-specific alternative splicing."
 RT J. Biol. Chem. 274:22053-22059 (1999).
 RL [3]
 CC -I- FUNCTION: Type XII collagen interacts with type I collagen-containing fibrils, the COL1 domain could be associated with the surface of the fibrils, and the COL2 and NC3 domains may be localized in the pericellular matrix (By similarity).
 CC -I- SUBUNIT: Trimer of identical chains each containing 190 kDa of nontriple-helical sequences (By similarity).
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=The final tissue form of collagen XII may contain homotrimers or any combination of the various isoforms;
 CC Name=XIIA-1;
 CC IsoId=Q60847-1; Sequence=Displayed;
 CC Name=XIIA-2; Synonyms=ER#K;
 CC IsoId=Q60847-2; Sequence=VSP_001151, VSP_001152;

CC Name=XIIB-1;
 CC IsoId=Q60847-3; Sequence=VSP_001150;
 CC Name=XIIB-2;
 CC IsoId=Q60847-4; Sequence=VSP_001150, VSP_001151, VSP_001152;
 CC TISSUE SPECIFICITY: Highest expression in tendons, perichondrium, skin, cornea, sclera, blood vessels, and peristomium.
 CC -I- DEVELOPMENTAL STAGE: The long NC3 XIIA isoforms are predominant at early stages (ED7 and 11); at later stages of development (ED15 and 17) the short NC3 XIIA forms become the major forms. As the short NC3 forms become the major product, the long splice variant continues to be expressed in several tissues, even after birth.
 CC The long NC1 isoforms, XIIA-1 and XIIA-1, peak in 15-day old embryos and decrease in 17-day old ones. The expression of the short NC1 form XIIA-2 remains constant throughout late stages of embryonic development (ED15 and ED17).
 CC -I- PTM: The triple-helical tail is stabilized by disulfide bonds at each end (By similarity).
 CC -I- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains (By similarity).
 CC -I- PTM: O-glycosylation of isoform XIIA-2; glycosaminoglycan of chondroitin-sulfate type (By similarity).
 CC -I- SIMILARITY: Belongs to the fibril-associated collagens with interrupted helices (FACIT) family.
 CC -I- SIMILARITY: Contains 18 fibronectin type-III domains.
 CC -I- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -I- SIMILARITY: Contains 4 WFPA domains.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
 CC -----
 CC EMBL: U55652; AAA9719.1; ALT SEQ. MENA.
 CC EMBL: U57095; AAB07047.1; -, mRNA.
 CC HSSP: P18614; IMHP.
 CC Ensembl: ENSMUSG00000032332; Mus musculus.
 CC GO: GO:0005615; C:extracellular space; TAS.
 CC InterPro: IPR008160; Collagen.
 CC InterPro: IPR003961; FN_III.
 CC InterPro: IPR003129; Laminin_G_TSP_N.
 CC InterPro: IPR002035; WVF_A.
 CC Pfam: PF01391; Collagen; 4.
 CC Pfam: PF00041; FN3; 18.
 CC Pfam: PF00092; WVF; 4.
 CC PRINTS: PRO0453; WVFADOMAIN.
 CC SMART: SM00060; FN3; 18.
 CC SMART: SM00210; TSPN; 1.
 CC SMART: SM00327; WVF; 4.
 CC PROSITE: PS50853; FN3; 18.
 CC PROSITE: PS50234; WVF; 4.
 CC Alternative splicing; Cell adhesion; Collagen; Extracellular matrix; Glycoprotein; Hydroxylation; Repeat; Signal; Structural protein.
 CC SIGNAL 1 24
 FT CHAIN 25 3119 Collagen alpha 1(XII) chain.
 FT DOMAIN 25 312 Fibronection type-III 1.
 FT DOMAIN 140 316 WVF 1.
 FT DOMAIN 333 422 Fibronection type-III 2.
 FT DOMAIN 444 620 WVF 2.
 FT DOMAIN 635 723 Fibronection type-III 3.
 FT DOMAIN 726 814 Fibronection type-III 4.
 FT DOMAIN 817 905 Fibronection type-III 5.
 FT DOMAIN 908 997 Fibronection type-III 6.
 FT DOMAIN 989 1087 Fibronection type-III 7.
 FT DOMAIN 1090 1179 Fibronection type-III 8.
 FT DOMAIN 1203 1375 WVF 3.
 FT DOMAIN 1388 1476 Fibronection type-III 9.
 FT DOMAIN 1478 1567 Fibronection type-III 10.
 FT DOMAIN 1569 1656 Fibronection type-III 11.
 FT DOMAIN 1660 1747 Fibronection type-III 12.
 FT DOMAIN 1758 1847 Fibronection type-III 13.

FT DOMAIN 1849 1937 Fibronectin type-III 14.
 FT DOMAIN 1939 2028 Fibronectin type-III 15.
 FT DOMAIN 2030 2119 Fibronectin type-III 16.
 FT DOMAIN 2121 2208 Fibronectin type-III 17.
 FT DOMAIN 2212 2298 Fibronectin type-III 18.
 FT DOMAIN 2329 2501 vWFA 4.
 FT DOMAIN 2525 2717 TSP N-terminal.
 FT REGION 2456 2751 Nonhelical region (NC3).
 FT REGION 2752 2899 Triple-helical region (COL2) with 1
 FT REGION 2900 2942 Nonhelical region (NC2).
 FT REGION 2943 3045 Triple-helical region (COL1) with 2
 FT REGION 3046 3119 Imperfections.
 FT MOTIF 866 868 Nonhelical region (NC1).
 FT MOTIF 2784 2786 Cell attachment site (Potential).
 FT MOTIF 2896 2898 Cell attachment site (Potential).
 FT COMBIAS 869 872 Poly-Thr.
 FT MOD_RES 2945 2945 Hydroxyproline (By similarity).
 FT MOD_RES 2948 2948 Hydroxyproline (By similarity).
 FT MOD_RES 2951 2951 Hydroxyproline (By similarity).
 FT MOD_RES 2960 2960 Hydroxyproline (By similarity).
 FT MOD_RES 2966 2966 Hydroxyproline (By similarity).
 FT MOD_RES 2969 2969 Hydroxyproline (By similarity).
 FT MOD_RES 2972 2972 Hydroxyproline (By similarity).
 FT MOD_RES 2984 2984 Hydroxyproline (By similarity).
 FT MOD_RES 3001 3001 Hydroxyproline (By similarity).
 FT MOD_RES 3004 3004 Hydroxyproline (By similarity).
 FT MOD_RES 3015 3015 Hydroxyproline (By similarity).
 FT MOD_RES 3024 3024 Hydroxyproline (By similarity).
 FT MOD_RES 3027 3027 Hydroxyproline (By similarity).
 FT MOD_RES 3030 3030 Hydroxyproline (By similarity).
 FT CARBOHYD 704 704 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 802 802 O-linked (Xyl...) (Chondroitin sulfate)
 FT CARBOHYD 893 893 O-linked (Xyl...) (Chondroitin sulfate)
 FT CARBOHYD 985 985 O-linked (Xyl...) (Chondroitin sulfate)
 FT CARBOHYD 1769 1769 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 2212 2212 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 2533 2533 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 2684 2684 N-linked (GlcNAc...) (Potential).
 FT VARSPLIC 25 1190 Missing (in isoform XIIB-1 and isoform
 XIIB-2).
 FT VARSPLIC 3062 3064 /FTId=VSP_001150.
 FT VARSPLIC 3065 3119 EPR -> GSG (in isoform XIIA-2 and isoform
 XIIB-2).
 FT VARSPLIC 3065 3119 /FTId=VSP_001151.
 FT VARSPLIC 3065 3119 Missing (in isoform XIIA-2 and isoform
 XIIB-2).
 FT VARSPLIC 3065 3119 /FTId=VSP_001152.
 FT VARSPLIC 3065 3119 9BIF959C86AB3251 CRC64;
 SQ SEQUENCE 3119 AA; 340244 MM; 9BIF959C86AB3251 CRC64;

Query Match 18.9%; Score 407; DB 1; Length 3119;
 Best Local Similarity 28.7%; Pred. No. 1.5e-18;
 Matches 121; Conservative 64; Mismatches 177; Indels 60; Gaps 10;

QY 23 ERGPPASAP-----RGDLMLFDSSASVSHYFGRVREFVQGLVAPLPGLGTG 69
 DB 420 EKTQPTQPMKQVQESRGVDIKADIVFLVDGYSIGTANFVKRAFLVLAASFELSPN 479
 QY 70 ALFASLVHVSRRPYTEPPFGOHSSGEAQAODAVASAOBMDTHTGLALVYAKOULFAEAS 129
 DB 480 RVOISLVQYSRDHTTEFTLKEPNRVEDITIKAINTFPRGSGSTNTGKMTYVRKIPVBNK 539
 QY 130 GARPVPKYLWVTGGSSDPVGPQGLKDLGVTVFVSTGRGNFLSLAASAPAEKH 189
 DB 540 GSRNVKPKWILITDGSDFAPDPAIKLNSDVEIPAVGVKDAVRELEAIAISPAAETH 599
 QY 190 LHFV-DVDDLIHIVQELRGI-----LDMR-----PQGLHATEITSSGFLAMPPL 235
 DB 600 VFTVEDFADAFQRISEFVLTSICLRIEQELAIKKKAVVPPKDLRFQVTANSFKAEMSP 659

QY 236 -----LRADSGTYLELVPSAQPGAAARQQLPGNATDWTIMAGLDPDTDYDVA 282
 DB 660 GNVPSHYTYXDANGDDEVTVVE-----PASTSVLMSLRRETLVLYVN 704
 QY 283 LVPESVRLRLRQILRVTRPREAG-PERIVISHARPRESLRVSWAPALGSAALGVHOF 341
 DB 705 VTAEEHEDGSPV-ITGEETTAEVKGVPRNLKVYDDETTDSFKLTWSQAPG--RVLYRIRY 761
 QY 342 GPLRGGEAQRVVEVPAQRNCTTQGLAPGTAVLVYVTAAPRSGRESALSAKACTPDGPRR 401
 DB 762 RVPSSGESKEVSTPANQRKTLLENLTPDKYELSVIAEYPSGPGSLTGNAAEE-VRGN 820
 QY 402 PR 403
 DB 821 PR 822

Search completed: February 13, 2006, 07:47:47
 Job time : 331.926 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
(c) 1993 - 2006 Biocore

Search time 254.435 Seconds

721.837 Million cell updates/sec

.....RPRPRVPRAPTPTGTASREP. 418

BLOSUM62

2443163

predicted by chance to have a score of the result being printed total score distribution.

TESTS

Description
Aae32502 Human Will
Abp69674 Human pol
Adh71106 Human pro
Aae32501 Mouse Wil
Aaw66326 Kidney in
Abd42581 Human ORF
Aab88340 Human mem
Ady65045 Human cld
Aab87344 Human gen
Abg65347 Human alb
Adl78614 Albunin f
Kab87418 Human gen
Aab87424 Human gen
Aae03654 Human ext
Aae32500 Human vor
Aae32503 Mouse vor
Abi47416 Breast ce
Adj75666 Marker ge
Abb90762 Human Tun
Abu54469 Human tun
Abi47415 Breast ce
Aau27790 Human fun
Adj75366 Marker ge
Adj65785 Human hee

[illegible]

ALIGNMENTS

RESULT 1
AAB32502
ID AAB32502 standard; protein; 418 AA.
XX
AC AAB32502;
XX
D7 24-MAR-2003 (first entry)
XX
DE Human Willebrand Factor A domain related-protein (
XX
XX Willebrand Factor A domain related-protein; von Willebrand
KW extracellular matrix; ECM; molecular marker; cartilage
XX gene therapy; human.
OS Homo sapiens.
PH Key Location/Qualifiers
FT Peptide 1..18
FT Protein /label= Signal_peptide
FT /note= "Human mature WARP protein"
FT Modified-site 148
FT /note= "O-glycosylation site"
FT Misc-difference 210..211
FT /note= "Encoded by CTCGGC"
FT Modified-site 264
FT /note= "N-glycosylation site"
FT Modified-site 359
FT /note= "N-glycosylation site"
FT Modified-site 361
FT /note= "O-glycosylation site"
FN Disulfide-bond 369..393
FN WO200288184-A1.
PD 07-NOV-2002.
PE 02-MAY-2002; 2002MO-AU000542.
PR 02-MAY-2001; 2001AU-0004701.
PA (MURD-) MURDOCH CHILDRENS RES INST.
PI Bateman JF, Fitzgerald DJ;
DR WP1: 2003-111873/10.
N-PSDB; AAD50399.

XX New isolated Willebrand Factor A-Related Protein polypeptide useful for
PT the manufacture of a medicament in the treatment of a disease condition
PT of the extracellular matrix, in particular arthritis.

XX Claim 11; Page 76-78; 103pp; English.

XX The invention relates to Willebrand Factor A domain related-protein
CC (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
CC superfamily of extracellular matrix (ECM) proteins. WARP is used as a
CC molecular marker, used for detecting a loss of ECM integrity in an animal
CC subject, monitoring repair, regeneration or other disease processes in an
CC animal subject and detecting a disease condition or a propensity for the
CC development of a disease condition in an animal subject. The invention is
CC useful for the manufacture of a medicament in the treatment of a disease
CC condition of the ECM. The disease condition involves the cartilage, and
CC is preferably arthritis. The invention is also used in gene therapy. The
CC present sequence is human WARP protein

XX Sequence 418 AA;

Query Match 100.0%; Score 2154; DB 6; Length 418;
Best Local Similarity 100.0%; Pred. No. 2.6e-179;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MLPTALGLALSLRLALRSAGRGPPASAPRGDLMFLDSSASVSHYEFRRVEFGQL 60
DB 1 MLPTALGLALSLRLALRSAGRGPPASAPRGDLMFLDSSASVSHYEFRRVEFGQL 60
QY 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQSSGGAADAVARASQRMGDTHTGALVYA 120
DB 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQSSGGAADAVARASQRMGDTHTGALVYA 120
QY 121 KEOLFPAEASGARPGVVKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 180
DB 121 KEOLFPAEASGARPGVVKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 180
QY 181 AASAPAEKHLHFVDDVDLHIIYQELRGSLIDAMRPOQLHATEITSSGFRFLAMPPLLTADS 240
DB 181 AASAPAEKHLHFVDDVDLHIIYQELRGSLIDAMRPOQLHATEITSSGFRFLAMPPLLTADS 240
QY 241 GYVVELVPSAOPGAARQOPLGNATDWTWAGLDPTDYVALVPSNVRLRLPOLILRVR 300
DB 241 GYVVELVPSAOPGAARQOPLGNATDWTWAGLDPTDYVALVPSNVRLRLPOLILRVR 300
QY 301 TPPEAGPERIVISHARPSRLVSNWAPALGSAALGYHGFQPLRGGEAQRVEVPAGRNC 360
DB 301 TPPEAGPERIVISHARPSRLVSNWAPALGSAALGYHGFQPLRGGEAQRVEVPAGRNC 360
QY 361 TTLQGLAPGTAVLVVTTAAFRSGRESALSAKACTPDGPRPRPRPVPRAPTGTASREP 418
DB 361 TTLQGLAPGTAVLVVTTAAFRSGRESALSAKACTPDGPRPRPRPVPRAPTGTASREP 418
```

RESULT 2
ABP69674
ID ABP69674 standard; protein; 445 AA.

XX ABP69674;

XX 20-JAN-2003 (first entry)
XX Human polypeptide SEQ ID NO 1721.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
XX cell-proliferative disorder; neurodegenerative disease; bacterial;
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;
XX multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
XX arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
XX antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
XX haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
XX antithratic.

OS Homo sapiens.

XX WO200270539-A2.

XX 12-SEP-2002.

XX 05-MAR-2002; 2002WO-US005095.

XX 05-MAR-2001; 2001US-00799451.

XX (HYSEQ-) HYSEQ INC.

XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
XX Wehrman T, Wang J, Wang D, Dimaanic RT;

XX WPI: 2002-759812/82.

XX N-PSDB; AB211891.

XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT or coagulation disorders.

XX Claim 9; SEQ ID NO 1721; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a
XX nucleotide sequence selected from any of 948 sequences (AB211119-
XX AB212066) or their mature protein coding portion, active domain coding
XX protein or complementary sequences. The polynucleotides are useful for
XX identifying expressed genes or for physical mapping of human genome. The
XX encoded polypeptides (ABP68902-ABP68849) are useful as molecular weight
XX markers, as a food supplement, for generating antibodies, in medical
XX imaging, screening and diagnostic assays and for treating cell-
XX proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
XX or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
XX diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
XX platelet or coagulation disorders, wound, burns, incision, ulcers, liver
XX or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
XX arthritis, etc. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 445 AA;

Query Match 98.9%; Score 2130.5; DB 5; Length 445;
Best Local Similarity 93.9%; Pred. No. 3.2e-177;
Matches 418; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

```
QY 1 MLPTALGLALSLRLALRSAGRGPPASAPRGDLMFLDSSASVSHYEFRRVEFGQL 60
DB 1 MLPTALGLALSLRLALRSAGRGPPASAPRGDLMFLDSSASVSHYEFRRVEFGQL 60
QY 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQSSGGAADAVARASQRMGDTHTGALVYA 120
DB 61 VAPLPLGTGALRASLVHVGSRPYTEFPFGQSSGGAADAVARASQRMGDTHTGALVYA 120
QY 121 KEOLFPAEASGARPGVVKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 180
DB 121 KEOLFPAEASGARPGVVKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSA 180
QY 181 AASAPAEKHLHFVDDVDLHIIYQELRGSLIDAMRPOQLHATEITSSGFRFLAMPPLLTADS 240
DB 181 AASAPAEKHLHFVDDVDLHIIYQELRGSLIDAMRPOQLHATEITSSGFRFLAMPPLLTADS 240
QY 241 GYVVELVPSAOPGAARQOPLGNATDWTWAGLDPTDYVALVPSNVRLRLPOLILRVR 300
DB 241 GYVVELVPSAOPGAARQOPLGNATDWTWAGLDPTDYVALVPSNVRLRLPOLILRVR 300
QY 301 TR-----PEEAGPERIVISHARPSRLVSNWAPALGSA 333
DB 301 TRPEAGPGASPGESGAPPTQLAALPAPEAGPERIVISHARPSRLVSNWAPALGSA 360
```


XX	Sequence 445 AA;	
SQ	Query Match Best Local Similarity 98.9%; Score 2130.5; DB 8; Length 445; Matches 418; Conservative 0; Mismatches 0; Indels 27; Gaps 1;	
QY	1 MLPTALGLIALSLRLALARSGBRGPASAPRGDLMFLDSSASVSHYEFSSRVREVGQL 60	
DB	1 MLPTALGLIALSLRLALARSGBRGPASAPRGDLMFLDSSASVSHYEFSSRVREVGQL 60	
QY	61 VAPLPLGTGALRASLVHVSRRPYTERPPGQSSGGAADAVASAGRMGDTHTGLALAYA 120	
DB	61 VAPLPLGTGALRASLVHVSRRPYTERPPGQSSGGAADAVASAGRMGDTHTGLALAYA 120	
QY	121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 180	
DB	121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 180	
QY	181 AASAPAEKHLHFVDDVDDHIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLITADS 240	
DB	181 AASAPAEKHLHFVDDVDDHIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLITADS 240	
QY	241 GYVLELVPSAQPGARROQLPGNATDWTWAGLDPTDVALVPSNVRLLRPQILRYR 300	
DB	241 GYVLELVPSAQPGARROQLPGNATDWTWAGLDPTDVALVPSNVRLLRPQILRYR 300	
QY	301 TR-----PEEAGPERIVISHARPSRLRVSWAPALGSAA 333	
DB	301 TRPEEAGPERIVISHARPSRLRVSWAPALGSAA 333	
QY	334 ALGYHVOFGPLRGGAEORVEVPAGRNCTTLOGIAPGTAVLVTTAAFRSGRESALSAAKAC 393	
DB	334 ALGYHVOFGPLRGGAEORVEVPAGRNCTTLOGIAPGTAVLVTTAAFRSGRESALSAAKAC 393	
QY	394 TPDGPRPRRPVPRAPTGTASREP 418	
DB	421 TPDGPRPRRPVPRAPTGTASREP 445	
RESULT 4	AAE32501 standard; protein; 415 AA.	
AC	AAE32501;	
DT	24-MAR-2003 (first entry)	
DE	Mouse Willebrand Factor A domain related-protein (WARP).	
KM	Willebrand Factor A domain related-protein; von Willebrand Factor A; VA; extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP; gene therapy; mouse.	
OS	Mus sp.	
XX	Key	Location/Qualifiers
XX	Peptide	1..18
XX	Protein	/label= Signal_Peptide
XX	Modified-site	19..415
XX	Modified-site	/note= "Mouse mature WARP protein"
XX	Modified-site	148
XX	Modified-site	/note= "O-glycosylation site"
XX	Modified-site	264
XX	Modified-site	/note= "N-glycosylation site"
XX	Modified-site	359
XX	Modified-site	/note= "N-glycosylation site"
XX	Modified-site	361
XX	Disulfide-bond	/note= "O-glycosylation site"
XX	Modified-site	369..393
XX	Modified-site	400
XX	Modified-site	/note= "O-glycosylation site"

PN	WO200288184-A1.	
XX		
PD	07-NOV-2002.	
XX		
PF	02-MAY-2002; 2002MO-AU000542.	
XX		
PR	02-MAY-2001; 2001AU-00004701.	
XX		
XX	(MURD-) MURDOCH CHILDRENS RES INST.	
PA	Bateman JF, Fitzgerald DJ;	
XX		
PI	WPI. 2003-111873/10.	
DR	N-PSDB; AAD50022, AAD50398.	
XX		
PT	New isolated Willebrand Factor A-Related Protein polypeptide useful for the manufacture of a medicament in the treatment of a disease	
PT	of the extracellular matrix, in particular arthritis.	
XX		
PS	Claim 13; Page 74-75; 103pp; English.	
XX		
CC	The invention relates to Willebrand Factor A domain related-protein (WARP) which is a member of von Willebrand Factor A (VA)-domain protein superfamily of extracellular matrix (ECM) proteins. WARP is used as an animal subject, monitoring repair, regeneration or other disease processes in an animal subject and detecting a disease condition or a propensity for the development of a disease condition in an animal subject. The invention is useful for the manufacture of a medicament in the treatment of a disease condition of the ECM. The disease condition involves the cartilage, and is preferably arthritis. The invention is also used in gene therapy. The present sequence is mouse WARP protein	
SQ	Sequence 415 AA;	
Query Match	76.8%; Score 1653.5; DB 6; Length 415;	
Best Local Similarity	77.5%; Pred. No. 1.4e-135;	
Matches	324; Conservative 32; Mismatches 59; Indels 3; Gaps 1;	
QY	1 MLPTALGLIALSLRLALARSGBRGPASAPRGDLMFLDSSASVSHYEFSSRVREVGQL 60	
DB	1 MLPTALGLIALSLRLALARSGBRGPASAPRGDLMFLDSSASVSHYEFSSRVREVGQL 60	
QY	61 VAPLPLGTGALRASLVHVSRRPYTERPPGQSSGGAADAVASAGRMGDTHTGLALAYA 120	
DB	61 VAPLPLGTGALRASLVHVSRRPYTERPPGQSSGGAADAVASAGRMGDTHTGLALAYA 120	
QY	121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 180	
DB	121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 180	
QY	181 AASAPAEKHLHFVDDVDDHIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLITADS 240	
DB	181 AASAPAEKHLHFVDDVDDHIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLITADS 240	
QY	241 GYVLELVPSAQPGARROQLPGNATDWTWAGLDPTDVALVPSNVRLLRPQILRYR 300	
DB	241 GYVLELVPSAQPGARROQLPGNATDWTWAGLDPTDVALVPSNVRLLRPQILRYR 300	
QY	301 TRPEEAGPERIVISHARPSRLRVSWAPALGSAAALGYHVOFGPLRGGAEORVEVPAGRNAC 360	
DB	301 TRPEEAGPERIVISHARPSRLRVSWAPALGSAAALGYHVOFGPLRGGAEORVEVPAGRNAC 360	
QY	361 TTLOGIAPGTAVLVTTAAFRSGRESALSAAKACTPDGPRPRRPVPRAPTGTASREP 418	
DB	361 TTLOGIAPGTAVLVTTAAFRSGRESALSAAKACTPDGPRPRRPVPRAPTGTASREP 415	
RESULT 5	AAW86326 standard; protein; 421 AA.	
ID	AAW86326	
XX	AAW86326;	

XX 01-MAR-1999 (first entry)
XX Kidney injury associated molecule HM059 protein.
DE Kidney injury associated molecule; kidney injury related molecule; KIM;
XX Kidney injury associated molecule; kidney injury related molecule; KIM;
XX tissue growth promotion; regeneration; renal condition;
XX acute renal failure; acute nephritis; tumour.
OS Rattus sp.
XX WO9853071-A1.
XX 26-NOV-1998.
XX 22-MAY-1998; 98WO-US010547.
XX 23-MAY-1997; 97US-0047490P.
XX 23-MAY-1997; 97US-0047491P.
XX (BIOJ) BIOGEN INC.
XX Sanicola-Nadel M, Hession CH, Wei H, Cate RL;
XX WPI; 1999-045312/04.
XX N-PSDB; AAV80611.
XX Kidney injury-associated molecule, KIM, polypeptides - upregulated in
XX injured or regenerating tissues, useful to promote tissue growth and
XX regeneration, especially to treat renal conditions.
XX Claim 17; Page 135-137; 213pp; English.
XX The present sequence represents a kidney injury associated molecule (KIM)
XX protein. KIM proteins can be administered therapeutically by expressing
XX KIM encoding polynucleotides, to promote growth and/or survival of
XX damaged tissue (e.g. renal tissue), since the KIM proteins are
XX upregulated in injured or regenerating (especially renal) tissues. KIM
XX fusion proteins, conjugates, antibodies and vectors can also be used
XX therapeutically, e.g. these or the KIM proteins may be included with an
XX acceptable carrier in pharmaceutical compositions, useful for therapy/
XX prophylaxis of conditions associated with dysfunction/dysregulation of
XX KIM genes or proteins, especially renal diseases or impairments of renal
XX function in humans (e.g. acute renal failure, acute nephritis). The
XX polynucleotides can be used to produce antisense sequences which, when
XX internalised into cells, can disrupt expression of a cellular KIM gene,
XX also useful in therapy (e.g. to block the growth of tumours dependent on
XX KIM for growth) or compositions. The proteins and polynucleotides are
XX useful diagnostically e.g. to detect and quantify renal injury/disease
XX (indicative of increased risk, or presence of, renal injury or impaired
XX function), or abnormal responses to tissue injury (indicative of
XX increased risk, or presence of, an autoimmune response or abnormal tissue
XX growth arising from/affecting renal tissue). The proteins can also be
XX used to locate KIM-producing cells (especially specific loci, e.g. tissue
XX masses abnormally producing/expressing KIM such as tumours arising
XX from/affecting renal tissue), by contacting cells with an imageable KIM-
XX binding reagent and imaging reagent accumulation
XX
SQ Sequence 421 AA;
Query Match 76.3%; Score 1643.5; DB 2; Length 421;
Best Local Similarity 77.8%; Pred. No. 1,1e-134;
Matches 325; Conservative 31; Mismatches 59; Indels 3; Gaps 1;
QY 1 MLEFWLGLALSLRLAARSGARPPAPAPRDLMFLDLSASVSHYEFRRVFGQL 60
DB 7 MLEFWLGLALSLRLAARSGARPPAPAPRDLMFLDLSASVSHYEFRRVFGQL 66
QY 61 VAPLPYGTGALRLSLVHVGSRPYTFPPGQSSGMAAOAVRARSARMDTHGTGLLVYA 120
DB 67 VAPLPYGTGALRLSLVHVGSRPYTFPPGQSSGMAAOAVRARSARMDTHGTGLLVYA 126
QY 121 KEQLFAEASGARPGVPRVLTWTDGSSDPVGPOMQLKDLGVTVEIVSTGRGNFLELSA 180

DB 127 KEQLFAEASGARPGVPRVLTWTDGSSDPVGPOMQLKDLGVTVEIVSTGRGNFLELSA 186
QY 181 AASAPAEKILHFDVDDHIIIVQELRGSITLDMRPOQLHATEITSSGFPLAMPPLTTADS 240
DB 187 AASAPAEKILHFDVDDHIIIVQELRGSITLDMRPOQLHATEITSSGFPLAMPPLTTADS 246
QY 241 GYVLELVPSAOPGAARQQLPGNATDWIMAGLDPTDVALVESNRYLLRPQILRVR 300
DB 247 GYVLELVPSAOPGAARQQLPGNATDWIMAGLDPTDVALVESNRYLLRPQILRVR 306
QY 301 TPPEAGPERIVISHTPRSLRVSAPALGSPSTLGYLVQLGSGSLHEVPAQONS 366
DB 307 TLQEEAGPERIVISHTPRSLRVSAPALGSPSTLGYLVQLGSGSLHEVPAQONS 366
QY 361 TLQEEAGPERIVISHTPRSLRVSAPALGSPSTLGYLVQLGSGSLHEVPAQONS 418
DB 367 TLQEEAGPERIVISHTPRSLRVSAPALGSPSTLGYLVQLGSGSLHEVPAQONS 421
RESULT 6
AAB42581
ID AAB42581 standard; protein; 299 AA.
XX AAB42581;
XX 08-FEB-2001 (first entry)
XX Human ORFX ORF2345 polypeptide sequence SEQ ID NO:4690.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
XX anticonvulsant; osteoplastic; antiarthritic; immunosuppressant; cardiac;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antineumatic; antihypertensive;
XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
XX 02-APR-1999; 99US-0127636P.
XX 05-APR-1999; 99US-0127728P.
XX 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimketa RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX N-PSDB; AAC76790.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 3880-3881; 5507pp; English.
XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiallergic; immunosuppressant; immunostimulant;
CC cadant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antineumatic; antithyroid; and antinaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

Sequence 299 AA;

Query Match 63.2%; Score 1362; DB 3; Length 299;
 Best Local Similarity 93.2%; Pred. No. 2.7e-110;
 Matches 274; Conservative 0; Mismatches 16; Indels 4; Gaps 2

QY	PMATLGAALSLRLALNRSQAGNERGPASARGMLFLDSSASVSHYEESRRREEVGQIVA	62
Db	PLDGARLUSLRM--RWRTPDCCPSABE-DIMFLDSSASVSHYEESRRVEEVGQIVA	65
QY	PLPLGTALRASLVHVGSRPYTEPPFGQSSGEAAQDVRASQCMGTHTGALVYAKE	122
Db	PLPLAPALRASLVHVGSRPYTEPFQGHSSGEAAQDVRASQCMGTHTGALVYAKE	125
QY	QLPFAASGAPGYPKVLWYTDGSSDPLGPPMOELKDLGVTVFVSTGRGNFLELSAA	182
Db	QLPFAASGAPGVPKVLWYTDGSSDPLGPPMOELKDLGVTVFVSTGRGNFLELSAAA	185
QY	SAPAEKLIHVDVDDHHIIYQELRGSIIIDAMPQOHLAETISSGFRILAMPPLITADSGY	242
Db	SAPAEKLIHVDVDDHHIIYQELRGSIIIDAMPQOHLAETISSGFRILAMPPLITADSGY	245
QY	YVLEIVSAPGGAARRQQLPGNATDWIMAGLDPTDIDVALVPESNVALLRPQI	296
Db	YVLEIVSAPGGAARRQQLPGNATDWIMAGLDPTDIDVALVPESNVALLRPQI	299

RESULT 7
AAB88340

ID AAB88340 standard; protein; 285 AA.

AC AAB88340;

DT 23-MAY-2001 (first entry)

DE Human membrane or secretory protein clone PSEC0053.

KW Human; secretory protein; membrane protein; vaccine; gene therapy, rheumatoid arthritis; diabetes.

PN EP1067182-A2.

PD 10-JAN-2001

07-JUL-2000; 2000EP-00114090.

PR 08-JUL-1999; 99JP-00194179

PR 11-JAN-2000; 2000JP-00118775

02-MAY-2000; 2000JP-00183766

PA (HELI -) HELIX RES INST.

PI Oca T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K,
XX
DR WPI, 2001-093589/11.
DR N-PSDB; AAF93167.

PT Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development.

PS Claim 1; SEQ ID NO 48; 609pp + Sequence Listing; English.

This invention relates to nucleic acid sequences AAF93744 - AAF9316 which encode human secretory or membrane proteins represented by AAB883177 - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the invention. The invention also includes methods for the production of antibodies directed against the proteins, and cDNA sequences, which can be used in vaccines. The polynucleotide sequences can be used in gene therapy. The polynucleotide sequences and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down-regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbent assay (ELISA). Examples of diseases which may be treated include rheumatoid arthritis and diabetes

Sequence 285 AA:

Query Match	59.8%	Score 1288.5;	DB 4;	Length 285;
Best Local Similarity	96.6%;	Pred. No. 6.7e-104;		
Matches 256; Conservative	2;	Mismatches 2;	Indels 5;	Gaps 1

QY	I	MLFWTLGLALSLRLALNASSGERRPPASAPRGDMLFLDSSASVHYEFSRRAREPFGOL	60
Db	1	MLFWTLGLALSLRLALNASSGERRPPASAPRGDMLFLDSSASVHYEFSRRAREPFGOL	60
QY	61	VAELPIGTGALNASTLVHNGSRPYTEFPFGOHSSGGAADAVRAASQRMGDTHTGALVYA	120
Db	61	VAELPIGTGALNASTLVHNGSRPYTEFPFGOHSSGGAADAVRAASQRMGDTHTGALVYA	120
QY	121	KEQLFNASGARRPGVPKVLVWTTDGGSSDPNGPPMQELKDLGTVVFVYVSGRGNFLELSA	180
Db	121	KEQLFNASGARRPGVPKVLVWTTDGGSSDPNGPPMQELKDLGTVVFVYVSGRGNFLELSA	180
QY	181	AASAPAEKHLHAFVDVDDLHIIYOELRGSLIDAMRFOQLHATEITSSSGRFLAMPPLPADS	240
Db	181	AASAPAEKHLHAFVDVDDLHIIYOELRGSLIDAMRFOQLHATEITSSSGRFLAMPPLPADS	240
QY	241	GYVLELVPSAQPG-----AARRQQ	260
Db	241	GYVLELVPSAQPGCKTPAARER	265

RESULT 8

ID ADY63045 standard; protein; 285 AA

AC ADY63045;

DT	02-JUN-2005 (first entry)
----	---------------------------

Human clone PSEC0053 protein, SEQ ID 48.

KW Gene therapy.

OS Homo sapiens.
XX
XX EP1514933-A1.
XX
XX 16-MAR-2005.
XX
PF 07-JUL-2000; 2004EP-00027228.
XX
XX 08-JUL-1999; 99JP-00194179.
XX 11-JAN-2000; 2000JP-00118775.
XX 02-MAY-2000; 2000JP-00183766.
XX 07-JUL-2000; 2000EP-00114090.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayaishi K;
XX
XX WPI; 2005-203865/22.
XX
XX N-PSDB; ADY63044.
XX
XX Novel isolated polynucleotide encoding human secretory proteins or
XX membrane proteins, useful for examination and diagnosis of abnormality of
XX human secretory proteins.
XX
XX Disclosure, SEQ ID NO 48; 1240bp; English.
XX
XX The present invention relates to novel human secretory proteins or
XX membrane proteins, and their coding sequences. The present sequence is
XX one such protein sequence. The coding sequences of the invention are
XX useful for examination and diagnosis of abnormality of the human
XX secretory proteins and in gene therapy methods. The coding sequences and
XX proteins are useful as candidates for medicines or as target molecules
XX for developing medicines. Antibodies against the proteins of the
XX invention are useful for treating diseases that are associated with the
XX proteins. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained from sequence information
XX supplied by the European Patent Office.
XX
XX Sequence 285 AA;
SQ
Query Match 59.8%; Score 1288.5; DB 9; Length 285;
Best Local Similarity 96.6%; Pred. No. 6.7e-104; Indels 5; Gaps 1;
Matches 256; Conservative 2; Mismatches 2;
QY 1 MLFWTALGLALSLRLALNARGRGPASAPRGDLMLFLDSSASVSHYEFRRVFEVQQL 60
1 MLFWTALGLALSLRLALNARGRGPASAPRGDLMLFLDSSASVSHYEFRRVFEVQQL 60
DB 1 MLFWTALGLALSLRLALNARGRGPASAPRGDLMLFLDSSASVSHYEFRRVFEVQQL 60
QY 61 VAPLPLGTGALRASLVHVGSRPYTFEPFGQSSGGEAAQDAVRASAOQRMGDTHTGLALVYA 120
61 VAPLPLGTGALRASLVHVGSRPYTFEPFGQSSGGEAAQDAVRASAOQRMGDTHTGLALVYA 120
DB 61 VAPLPLGTGALRASLVHVGSRPYTFEPFGQSSGGEAAQDAVRASAOQRMGDTHTGLALVYA 120
QY 121 KEQLFPAASGARCVKRVLMWTDGSSDPVGPMPQELKDLGTVTFIVSGRNFLELSA 180
121 KEQLFPAASGARCVKRVLMWTDGSSDPVGPMPQELKDLGTVTFIVSGRNFLELSA 180
DB 121 KEQLFPAASGARCVKRVLMWTDGSSDPVGPMPQELKDLGTVTFIVSGRNFLELSA 180
QY 181 AASAPAEKHLHFDVVDLHITVOELRGSIIDARPOOLHATEITSSGFRLLAPPLLTADS 240
181 AASAPAEKHLHFDVVDLHITVOELRGSIIDARPOOLHATEITSSGFRLLAPPLLTADS 240
DB 181 AASAPAEKHLHFDVVDLHITVOELRGSIIDARPOOLHATEITSSGFRLLAPPLLTADS 240
QY 241 GYVVLGLVPSAQPGCGCTPAAAMER 265
241 GYVVLGLVPSAQPGCGCTPAAAMER 265
DB 241 GYVVLGLVPSAQPGCGCTPAAAMER 265
RESULT 9
AAB87344 ID AAB87344 standard; protein; 215 AA.
XX
XX AAB87344;
AC
XX
XX 22-MAY-2001 (first entry)
DT
XX

DE Human gene 3 encoded secreted protein HNTB078, SEQ ID NO:85.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
XX cardiovascular disorder; angiogenic disorder; kidney disorder;
XX gastrointestinal disorder; pregnancy-related disorder;
XX endocrine disorder; infection; wound healing; vlnerrary; cell culture;
XX chemotaxis; food additive; binding partner identification.
XX
XX Homo sapiens.
XX
XX WO200118022-A1.
XX
XX 15-MAR-2001.
XX
XX 31-AUG-2000; 2000MO-US024008.
XX
XX 03-SEP-1999; 99US-0152315P.
XX
XX 03-SEP-1999; 99US-0152317P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Baker KP, Birse CE, Fiscella M, Komatsuolis GA, Rosen CA;
XX Soppe DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
XX Moore PA, Shi Y, Wei Y, Florence KA;
XX
XX WPI; 2001-203081/20.
XX
XX N-PSDB; AAF91860.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
XX Parkinson's diseases and cancers.
XX
XX Claim 11; Page 532-533; 607bp; English.
XX
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
XX protein genes, and AAB87342-AAB87413 represent the proteins they encode.
XX AAB87414-AAB87454 represent human secreted protein fragments. The genes
XX and their corresponding secreted proteins are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 52 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
XX disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX cardiovascular disorders, angiogenic disorders, kidney disorders,
XX gastrointestinal disorders, pregnancy-related disorders, endocrine
XX disorders, and infections. The proteins can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues, to identify their
XX cognate ligands or binding partners, and in chemotaxis, and can be used
XX as a food additive or preservative to modify storage properties.
XX Antibodies specific for a protein of the invention can be used in
XX alleviating symptoms associated with the disorders mentioned above, and
XX in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX immunosorbent assay (ELISA). The present sequence represents a human
XX secreted protein of the invention
XX
XX Sequence 215 AA;
SQ
Query Match 50.3%; Score 1084; DB 4; Length 215;
Best Local Similarity 99.5%; Pred. No. 3.4e-86;
XX

	Matches	214;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0
Qy		1	MLPFWALGIALSLRLALRSGAERGPPASAPGDMLFLIDSSASVSHYFSRVERFVQL	60						
Db		1	MLPFWAXGIALSLRLALRSGAERGPPASAPGDMLFLIDSSASVSHYFSRVERVQL	60						
Qy		61	VAPFLGTGALPASLVHVGSRRYTEPPFGQSHSGEAAQDAVPASQRMGDTHGLALVYA	120						
Db		61	VAPFLGTGALPASLVHVGSRRYTEPPFGQSHSGEAAQDAVPASQRMGDTHGLALVYA	120						
Qy		121	KEQLFAEASGARPGPKVLVWVTDGSSDPVCPMPQELDLGVTVIVSTGKGNFLEISA	180						
Db		121	KEQLFAEASGARPGPKVLVWVTDGSSDPVCPMPQELDLGVTVIVSTGKGNFLEISA	180						
Qy		181	AASAPAEKLIHFVDVDDLHIIVQELRGSLIDLMRP	215						
Db		181	AASAPAEKLIHFVDVDDLHIIVQELRGSLIDLMRP	215						
RESULT 10										
ABG65347										
ID			ABG65347	standard; protein; 215 AA.						
XX			ABG65347;							
AC										
XX										
DT			27-AUG-2002	(first entry)						
XX										
DE				Human albumin fusion protein #2022.						
KM				Albumin fusion protein; therapeutic protein X; human albumin; HA;						
KM				human serum albumin; HSA; cancer; reproductive disorder;						
KM				digestive disorder; immune disorder; endocrine disorder;						
KM				haematopoietic disorder; neural disorder; connective disorder;						
KM				cytosolic; antifertility; antiinflammatory; anticancer;						
KM				immunomodulator; anti-HIV; antidiabetic; haemostatic; nocrotropic;						
KM				neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;						
XX				osteopathic; antiarthritic.						
XX										
OS				Homo sapiens.						
XX				Synthetic.						
PN				WO200177137-A1.						
PD				18-OCT-2001.						
XX										
PF				12-APR-2001; 2001WO-US011988.						
XX										
PR				12-APR-2000; 2000US-0229358P.						
PR				25-APR-2000; 2000US-0199384P.						
XX										
XX				21-DEC-2000; 2000US-0256931P.						
XX										
PA				(HUMA-) HUMAN GENOME SCI INC.						
XX										
PI				Rosen CA, Haeeltime WA;						
XX										
DR				WPI; 2002-010886/01.						
PT										
XX				New fusion protein for treating disease e.g. diabetes comprises an						
XX				albumin fused to a therapeutic protein.						
PS				Claim 1; Page 1935; 2102pp; English.						
CC										
CC				The present invention relates to albumin fusion proteins comprising a						
CC				therapeutic protein X and human albumin (HA, also known as human serum						
CC				albumin, HSA). The proteins are useful for treating a disease or disorder						
CC				that may be modulated by therapeutic protein X. The albumin extends the						
CC				shelf-life of protein X, and may increase its biological in vitro/in vivo						
CC				activity. The protein X, and may increase its biological in vitro/in vivo						
CC				such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's						
CC				disease, ulcerative colitis), immune disorders (e.g. acquired						
CC				immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),						
CC				haematopoietic disorders, neural disorders (e.g. Alzheimer's,						
CC				Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,						

Query Match	50.3%;	Score 1084;	DB 5;	Length 215;
Best Local Similarity	99.5%;	Pred. No. 3,4e-86;		
Matches 214;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MLPWTALGLALSLRLALASGAEGRPPASAPRGDLMFLDSSASVSHYFSGRVAEFPGQL	60	
Db	1	MLPWTAXGLALSLRLALASGAEGRPPASAPRGDLMFLDSSASVSHYFSGRVAEFPGQL	60	
Qy	61	VAPLPFGTGLARASLVHVGSRPYTEFPFGQHSGBAAQDAVRASAKRMGDTHTGLALVYA	120	
Db	61	VAPLPFGTGLARASLVHVGSRPYTEFPFGQHSGBAAQDAVRASAKRMGDTHTGLALVYA	120	
Qy	121	KEQLFPAASARGGVKVLWVTTDGGSSDPVGPQDELKDIGVYFVYSTGRGNFLELSA	180	
Db	121	KEQLFPAASARGGVKVLWVTTDGGSSDPVGPQDELKDIGVYFVYSTGRGNFLELSA	180	
Qy	181	AASAPAEKHLHFDVDDHLITVOELRGSLDAMP	215	
Db	181	AASAPAEKHLHFDVDDHLITVOELRGSLDAMP	215	
RESULT 11				
ADL78614				
ID	ADL78614	standard; protein; 215 AA.		
AC	ADL78614;			
XX				
DT	20-MAY-2004	(first entry)		
XX				
DE	Albumin fusion protein related therapeutic protein X, SEQ ID NO 2096.			
XX				
KM	albumin fusion protein; cytostatic; anti-naemic; anti-arthritic;			
KM	antidiabetic; anti-HIV; immunosuppressive; anti-inflammatory;			
KM	antibacterial; osteopathic; dermatological; angiot;			
KM	immunomodulator; antiarrhythmic; cardiac; nootropic; analgesic;			
KM	neurotropic; unoparic; neuroprotective; antiparkinsonian; tranquilizer			
KM	antidiabetic; anabolic; hypertensive; vulnerary; gene therapy; cancer;			
KM	reproductive system disorder; therapeutic protein.			
XX				
OS	Unidentified.			
XX				
PN	US2004010134-A1.			
XX				
PD	15-JAN-2004.			
XX				
PF	12-APR-2001; 2001US-00833245.			
XX				
PR	12-APR-2000; 2000US-0229358P.			
PR	25-APR-2000; 2000US-0199384P.			
PR	21-DEC-2000; 2000US-0256931P.			
XX				
PA	(ROSE/) ROSEN C A.			
PA	(HASE/) HASELTINE W A.			
XX				
PI	Rosen CA, Haseltine WA;			
XX				
DR	WPI; 2004-090519/09.			
XX				
PT	New albumin fusion proteins, useful for diagnosing, treating, preventing			
PT	or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,			
PT	asthma, inflammatory bowel disease or Alzheimer's disease.			
XX				
PS	Disclosure; SEQ ID NO 2096; 279bp; English.			
CC	The invention relates to a novel albumin fusion protein. The invention			
CC	further relates to: a composition comprising the albumin fusion protein			
CC	and a pharmaceutical carrier; a kit comprising the albumin fusion protein			
CC	albumin fusion protein formula; a method of treating a disease or			

CC disorder in a patient comprising the step of administering the albumin
CC fusion protein; a method of treating a patient with a disease or disorder
CC that is modulated by Therapeutic Protein: X, or its fragment or variant;
CC a method of extending the shelf life of Therapeutic Protein: X, or its
CC fragment or variant; a nucleic acid molecule comprising a polynucleotide
CC sequence encoding the albumin fusion protein; a vector comprising the
CC nucleic acid molecule of the albumin fusion protein; and a host cell
CC comprising the nucleic acid molecule of the albumin fusion protein. The
CC albumin fusion protein and its compositions have the following
CC activities: cyostatic, antianemic, antiarthritic, antiaesthetic, anti-
CC HIV, immunosuppressive, antiinflammatory, antiproliferative, antibacterial,
CC osteopathic, dermatological, antigout, immunomodulator, antiarrhythmic,
CC cardiant, nootropic, antilipemic, nephrotropic, uropathic,
CC neuroprotective, antiparinsonian, tranquilizer, antidiabetic, anabolic,
CC hypertensive, and vulnerary. The albumin fusion protein nucleic acid may
CC be used in gene therapy to treat disorders. The albumin fusion protein is
CC useful for diagnosing, treating, preventing or ameliorating diseases or
CC disorders comprising: indication: X. The diseases or disorders include:
CC cancer (e.g. leukemia, colon, bone, breast, liver or lung cancer),
CC immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute
CC lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS,
CC autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
CC disease), reproductive system disorders (e.g. prostatitis, inguinal
CC hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
CC Leydig tumours), musculoskeletal diseases (e.g. giant cell tumours,
CC Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy
CC or cachexia), cardiovascular disease (e.g. rhadomyoma, heart disease,
CC arrhythmia, cardiac arrest, heat valve disease, hypernatremia or
CC hyponatremia), mixed foetal diseases (e.g. foetal alcohol syndrome,
CC Down's syndrome, Patou syndrome, Turner's syndrome, Apert syndrome or Tay-
CC Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
CC tract infections or renal disorders), neural or sensory disease (e.g.
CC Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
CC cerebellar ataxia, attention deficit disorder, autism or obsessive
CC compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or
CC occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
CC disease or glomerulonephritis), digestive diseases (e.g. portal
CC hypertension, irritable bowel disease, gastric atrophy or pancreatitis)
CC or connective tissue or epithelial diseases (e.g. Crohn's disease,
CC scleroderma, wound healing or epidermolysis bullosa). This sequence
CC represents a therapeutic protein X relating to the albumin fusion protein
CC of the invention. The sequence listing data for this specification was
CC downloaded from the USPTO website.
XX
XX
SQ Sequence 215 AA;
Query Match 50.3%; Score 1084; DB 8; Length 215;
Best Local Similarity 99.5%; Pred. No. 3,4e-86;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLPWTALGLALSRLLALSGAERGPPAPRDGLFLDSSASVGHYEFRRREFVQGL 60
DB 1 MLPWTALGLALSRLLALSGAERGPPAPRDGLFLDSSASVGHYEFRRREFVQGL 60
QY 61 VAPLPGLTGLRALSLVHVGSRPYTEPPFGHSSGEAAQDAVRASAOQMDTHTGLALVYA 120
DB 61 VAPLPGLTGLRALSLVHVGSRPYTEPPFGHSSGEAAQDAVRASAOQMDTHTGLALVYA 120
QY 121 KEPLFPAASGARGCVKVLWMTDGSDDVGPMPMELKDLGTVTFVSTGRNPFELSA 180
DB 121 KEPLFPAASGARGCVKVLWMTDGSDDVGPMPMELKDLGTVTFVSTGRNPFELSA 180
QY 181 AASAPAEKHLHFVDVDDLHITVDELRSIILNAMP 215
DB 181 AASAPAEKHLHFVDVDDLHITVDELRSIILNAMP 215
RESULT 12
ID AAB87418 standard; protein; 242 AA.
XX
XX AAB87418;
XX

DT 22-MAY-2001 (first entry)
XX Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:159.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnerary; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
XX Homo sapiens.
OS
XX
XX MO200118022-A1.
PD 15-MAR-2001.
PD
XX 31-AUG-2000; 2000WO-US024008.
PF
XX 03-SRP-1999; 99US-0152315P.
PR 03-SRP-1999; 99US-0152317P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Baker KP, Birse CE, Fiscella M, Komatsu GA, Rosen CA;
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
XX WPI; 2001-203081/20.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
XX
XX Disclosure; Page 18; 607p; English.
XX
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunoassay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX
XX
SQ Sequence 242 AA;
Query Match 50.3%; Score 1084; DB 4; Length 242;
XX

Best Local Similarity 99.5%; Pred. No. 4e-86;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLPWTALGLALSLRLALRSGAERGPPASAPRGDMLFLDSSASVSHYEFSSVREFFVQOL 60
DB 28 MLPWTALGLALSLRLALRSGAERGPPASAPRGDMLFLDSSASVSHYEFSSVREFFVQOL 87
QY 61 VAPPLGTCALRASLVHGSRYTEFFPGQSSGGAADAVARASQRMGDTHGTALVYA 120
DB 88 VAPPLGTCALRASLVHGSRYTEFFPGQSSGGAADAVARASQRMGDTHGTALVYA 147
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 180
DB 148 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 207
QY 181 AASAPAEKHLHFVDVDDHLITVQELRGSILDMARP 215
DB 208 AASAPAEKHLHFVDVDDHLITVQELRGSILDMARP 242

RESULT 13
AAB87424
ID AAB87424 standard; protein; 226 AA.
XX
AC AAB87424;
XX
DT 22-MAY-2001 (first entry)
XX
DE Human gene 3 encoded secreted protein fragment, SEQ ID NO:165.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery; cell culture;
KW chemotaxis; food additive; binding partner identification.
XX
OS Homo sapiens.
XX
PN WO200118022-A1.
XX
PD 15-MAR-2001.
XX
PE 31-AUG-2000; 2000MO-US024008.
XX
PR 03-SEP-1999; 99US-0152315P.
PR 03-SEP-1999; 99US-0152317P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsu G, Rosen CA,
PI Soppe DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW,
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
DR WPI; 2001-203081/20.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
PS Disclosure; Page 18; 607pp; English.
XX
CC AAP91858-AAP91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of

CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention
XX

SO Sequence 226 AA;

Query Match 50.0%; Score 1077; DB 4; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.5e-85;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPWTALGLALSLRLALRSGAERGPPASAPRGDMLFLDSSASVSHYEFSSVREFFVQOL 60
DB 14 MLPWTALGLALSLRLALRSGAERGPPASAPRGDMLFLDSSASVSHYEFSSVREFFVQOL 73
QY 61 VAPPLGTCALRASLVHGSRYTEFFPGQSSGGAADAVARASQRMGDTHGTALVYA 120
DB 74 VAPPLGTCALRASLVHGSRYTEFFPGQSSGGAADAVARASQRMGDTHGTALVYA 133
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 180
DB 134 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGKGFLELSA 193
QY 181 AASAPAEKHLHFVDVDDHLITVQELRGSILDMARP 213
DB 194 AASAPAEKHLHFVDVDDHLITVQELRGSILDMARP 226

RESULT 14
AAB03654
ID AAB03654 standard; protein; 185 AA.
XX
AC AAB03654;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human extracellular matrix and cell adhesion molecule-18 (XNAD-18).
XX
KW Human; extracellular matrix and cell adhesion molecule; XNAD;
KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
KW Down's syndrome; cystic fibrosis; Gaucher's disease; myoclonic dystrophy;
KW sickle cell anaemia; thalassemia; autoimmune disorder; adenocarcinoma;
KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
KW infection; cell proliferative disorder; actinic keratosis; myeloma;
KW arteriosclerosis; neotropic; anticonvulsant; antithyroid; nephrotropic;
KW neuroprotective; dermatological.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Region 1..171

FT FT /note= "Collagen glycoprotein precursor"
 FT Region 1. .170
 FT /note= "Von Willebrand factor domain score"
 FT Region 2. .15
 FT /note= "Collagen glycoprotein precursor"
 FT Region 37. .51
 FT /note= "Collagen glycoprotein precursor"
 FT Region 103. .111
 FT /note= "Collagen glycoprotein precursor"
 PN MO200142285-A2.
 PD 14-JUN-2001.
 PF 05-DEC-2000; 2000WO-US032990.
 XX
 XX
 PR 10-DEC-1999; 99US-0172852P.
 PR 16-DEC-1999; 99US-0172354P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PI Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
 PI Baughn MR, Lu DAM, Shah P, Au-Young J;
 XX WPI; 2001-381632/40.
 DR N-PSDB; AAD08062.
 DR
 PT New human extracellular matrix and cell adhesion molecules and
 PT polynucleotide sequences encoding them, useful for diagnosis, prevention,
 PT treatment of genetic, autoimmune and cell proliferative disorders.
 XX
 XX Claim 1, Page 111, 135pp; English.
 XX
 CC The present sequence is a human extracellular matrix and cell adhesion
 CC molecule (XMAP). The XMAP is used for screening a compound for
 CC effectiveness as an agonist or antagonist of XMAP. The identified agonist
 CC or antagonist are used for treating a disease or condition associated
 CC with decreased or increased expression of functional XMAP. The
 CC polynucleotides encoding XMAP are useful in somatic or germline gene
 CC therapy to correct a genetic deficiency, to express a conditionally
 CC lethal gene product and to express a protein which affords protection
 CC against intracellular parasites and also for diagnosis of disorders
 CC associated with expression of XMAP. They are also used for generating
 CC hybridization probes useful in mapping the naturally occurring genomic
 CC sequences and to create knock in humanised animals (pigs) or transgenic
 CC animals (mice or rats) to model human diseases. Oligonucleotide or longer
 CC fragments derived from the polynucleotide sequences may be used as
 CC elements on a microarray. Antibodies which specifically bind XMAP may be
 CC used for the diagnosis of disorders associated with the expression of
 CC XMAP, or in assays to monitor patients being treated with XMAP. Diseases
 CC diagnosed, prevented or treated include genetic disorders such as
 CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 CC disease, myotonic dystrophy, sickle cell anaemia, thalassemia,
 CC autoimmune/inflammatory disorders such as acquired immune deficiency
 CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
 CC atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
 CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 CC bacterial, fungal, parasitic, protozoal and helminthic infections and
 CC cell proliferative disorders such as actinic keratosis, arteriosclerosis
 CC and cancer including breast, bladder, bone marrow, brain and uterus
 CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma
 XX
 XX Sequence 165 AA;
 SO
 QY Query Match 42.6%; Score 918; DB 4; Length 185;
 QY Best Local Similarity 100.0%; Pred. No. 8, 9e-72;
 QY Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 36 MFLLDSSASVSHSEFRVREFGVQLVAPLPLGALRASLVHGRSPYTFEPGQSSGE 95
 DB 1 MFLLDSSASVSHSEFRVREFGVQLVAPLPLGALRASLVHGRSPYTFEPGQSSGE 60

QY 96 AAQDAVRASAGMGSTHTGLALVYAKEQLFAFASGARPCVPKYLWWTDDGSSDPVGPMM 155
 DB 61 AAQDAVRASAGMGSTHTGLALVYAKEQLFAFASGARPCVPKYLWWTDDGSSDPVGPMM 120
 QY 156 QELKDLGTVTFIVSGRGFELSAASAAPAKKHLPFVVDVDDLHIIVQELRGSIIIDAMP 215
 DB 121 QELKDLGTVTFIVSGRGFELSAASAAPAKKHLPFVVDVDDLHIIVQELRGSIIIDAMP 180
 QY 216 Q 216
 DB 181 Q 181
 RESULT 15
 AAE32500
 ID AAE32500 strand; protein; 180 AA.
 XX
 AC AAE32500;
 DT 24-MAR-2003 (first entry)
 XX
 DE Human von Willebrand Factor A (VA) domain.
 XX
 KW Willebrand Factor A domain related-protein; von Willebrand Factor A; VA;
 KW extracellular matrix; ECM; molecular marker; cartilage; arthritis; WARP;
 KW gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1. .2
 FT /note= "Encoded by GGG"
 FT Misc-difference 179. .180
 FT /note= "Encoded by CTC"
 PN MO200288184-A1.
 XX
 PD 07-NOV-2002.
 PF 02-MAY-2002; 2002WO-AU000542.
 XX
 PR 02-MAY-2001; 2001AU-00004701.
 PA (MURD-) MURDOCH CHILDRENS RES INST.
 PI Bateman JF, Fitzgerald DJ;
 XX WPI; 2003-111873/10.
 DR N-PSDB; AAD50397.
 PT New isolated Willebrand Factor A-Related Protein polypeptide useful for
 PT the manufacture of a medicament in the treatment of a disease condition
 PT of the extracellular matrix, in particular arthritis.
 XX
 PS Claim 7, Page 72-73; 103pp; English.
 XX
 CC The invention relates to Willebrand Factor A domain related-protein
 CC (WARP) which is a member of von Willebrand Factor A (VA)-domain protein
 CC superfamily of extracellular matrix (ECM) proteins. WARP is used as a
 CC molecular marker, used for detecting a loss of ECM integrity in an animal
 CC subject, monitoring repair, regeneration or other disease processes in an
 CC animal subject and detecting a disease condition or a propensity for the
 CC development of a disease condition in an animal subject. The invention is
 CC useful for the manufacture of a medicament in the treatment of a disease
 CC condition of the ECM. The disease condition involves the cartilage, and
 CC is preferably arthritis. The invention is also used in gene therapy. The
 CC present sequence is human VA domain
 XX
 XX Sequence 180 AA;
 SO
 QY Query Match 42.4%; Score 913; DB 6; Length 180;
 QY Best Local Similarity 100.0%; Pred. No. 2, 3e-71;
 QY Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 13:10:29 ; Search time 85.2776 Seconds
(without alignments)
2048.049 Million cell updates/sec

Title: US-10-699-035A-6

Perfect score: 2154
Sequence: 1 MLPWTALGLALSRLALARS.....RPRPRPVPAPPTGTASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:*
1: /cgn2_6/prodata/1/pubpaa/us07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpaa/us08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubpaa/us09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubpaa/us10_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpaa/us10_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubpaa/us11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2154	100.0	418	US-10-699-035A-6	Sequence 6, App1
2	2154	100.0	418	US-10-699-035A-20	Sequence 20, App1
3	1653.5	76.8	415	US-10-699-035A-4	Sequence 4, App1
4	1653.5	76.8	415	US-10-699-035A-21	Sequence 21, App1
5	1084	50.3	215	US-09-789-561-85	Sequence 85, App1
6	1084	50.3	215	US-09-833-245-2096	Sequence 2096, App1
7	1084	50.3	215	US-10-883-936-85	Sequence 85, App1
8	1084	50.3	242	US-09-789-561-159	Sequence 159, App1
9	1084	50.3	242	US-10-883-936-159	Sequence 159, App1
10	1077	50.0	226	US-09-789-561-165	Sequence 165, App1
11	1077	50.0	226	US-10-883-936-165	Sequence 165, App1
12	918	42.6	185	US-10-149-819-18	Sequence 18, App1
13	913	42.4	180	US-10-699-035A-2	Sequence 2, App1
14	772	35.7	180	US-10-699-035A-31	Sequence 31, App1
15	768	35.8	181	US-10-699-035A-8	Sequence 8, App1
16	424	19.7	3063	US-10-177-293-63	Sequence 63, App1
17	424	19.7	3063	US-10-631-467-918	Sequence 918, App1
18	421.5	19.6	3063	US-09-918-715-257	Sequence 257, App1
19	421.5	19.6	3063	US-10-177-293-61	Sequence 61, App1
20	421.5	19.6	3063	US-10-301-822-26	Sequence 26, App1
21	421.5	19.6	3063	US-10-474-794-257	Sequence 257, App1
22	421.5	19.6	3063	US-10-979-159-257	Sequence 257, App1
23	421.5	19.6	3118	US-10-320-335-287	Sequence 287, App1
24	407	18.9	3067	US-10-631-467-1618	Sequence 1618, App1
25	392.5	18.2	1207	US-10-408-765A-1591	Sequence 1591, App1
26	392.5	18.2	1297	US-10-187-975-102	Sequence 102, App1
27	382.5	17.8	2502	US-10-741-600-1627	Sequence 1627, App1

28	382.5	17.8	2944	US-10-723-860-1785	Sequence 1785, App1
29	382.5	17.8	2944	US-10-741-600-1629	Sequence 1629, App1
30	366	17.0	87	US-10-699-035A-40	Sequence 40, App1
31	365	16.9	87	US-10-699-035A-36	Sequence 36, App1
32	360	16.7	1899	US-10-631-467-919	Sequence 919, App1
33	326	15.1	1780	US-10-115-479-10	Sequence 10, App1
34	272.5	12.7	176	US-09-976-782-38	Sequence 38, App1
35	268.5	12.5	843	US-10-408-765A-17	Sequence 17, App1
36	265.5	12.3	182	US-10-699-035A-22	Sequence 22, App1
37	264	12.3	755	US-09-919-497-57	Sequence 57, App1
38	263.5	12.2	184	US-10-699-035A-27	Sequence 27, App1
39	259	12.0	580	US-10-723-860-2769	Sequence 2769, App1
40	253.5	11.8	776	US-10-000-512-8	Sequence 8, App1
41	253.5	11.8	776	US-10-074-566-8	Sequence 8, App1
42	253.5	11.8	782	US-10-428-275-176	Sequence 176, App1
43	253.5	11.8	794	US-09-833-245-1292	Sequence 1292, App1
44	253.5	11.8	914	US-10-428-275-156	Sequence 156, App1
45	253.5	11.8	915	US-09-909-320-34	Sequence 34, App1

ALIGNMENTS

RESULT 1					
US-10-699-035A-6					
; Sequence 6, Application US/10699035A					
; Publication No. US20040214349A1					
GENERAL INFORMATION:					
; APPLICANT: Bateman, John					
; TITLE OF INVENTION: A Molecular Marker					
; FILE REFERENCE: A36056 PCT USA A 071838.0142					
; CURRENT APPLICATION NUMBER: US/10/699,035A					
; PRIOR FILING DATE: 2003-10-31					
; PRIOR APPLICATION NUMBER: PCT/AU02/00542					
; PRIOR FILING DATE: 2002-05-02					
; PRIOR APPLICATION NUMBER: AU PR4701/01					
; PRIOR FILING DATE: 2001-05-02					
; NUMBER OF SEQ ID NOS: 40					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 6					
; LENGTH: 418					
; TYPE: PRT					
; ORGANISM: Homo sapiens					
US-10-699-035A-6					
Query Match					
Best Local Similarity 100.0%; Score 2154; DB 4; Length 418;					
Matches 418; Conservative 0; Pred. No. 2e-174; Indels 0; Gaps 0;					
Matches 418; Mismatches 0;					
QY	1	MLPWTALGLALSRLALARS	GAPPPAPRGDLMFLDSSASVSHYFRRVFGOL	60	
DB	1	MLPWTALGLALSRLALARS	GAPPPAPRGDLMFLDSSASVSHYFRRVFGOL	60	
QY	61	VAPPLGTGALPASLVHVS	RPYTFEPFGQSSGGAQAQAVASAGRMGDHTGALVYA	120	
DB	61	VAPPLGTGALPASLVHVS	RPYTFEPFGQSSGGAQAQAVASAGRMGDHTGALVYA	120	
QY	61	VAPPLGTGALPASLVHVS	RPYTFEPFGQSSGGAQAQAVASAGRMGDHTGALVYA	120	
DB	61	VAPPLGTGALPASLVHVS	RPYTFEPFGQSSGGAQAQAVASAGRMGDHTGALVYA	120	
QY	121	KEQLFASAGARPVKXUL	VWTTDGGSDPVGPMKELDGLGVTVITVSTGKGFLELA	180	
DB	121	KEQLFASAGARPVKXUL	VWTTDGGSDPVGPMKELDGLGVTVITVSTGKGFLELA	180	
QY	121	KEQLFASAGARPVKXUL	VWTTDGGSDPVGPMKELDGLGVTVITVSTGKGFLELA	180	
DB	121	KEQLFASAGARPVKXUL	VWTTDGGSDPVGPMKELDGLGVTVITVSTGKGFLELA	180	
QY	181	AASAPAEKILHFVDVDD	HLIIYOELRGSILDMRPOQLATETITSSGFRLAMPPLTTADS	240	
DB	181	AASAPAEKILHFVDVDD	HLIIYOELRGSILDMRPOQLATETITSSGFRLAMPPLTTADS	240	
QY	181	AASAPAEKILHFVDVDD	HLIIYOELRGSILDMRPOQLATETITSSGFRLAMPPLTTADS	240	
DB	181	AASAPAEKILHFVDVDD	HLIIYOELRGSILDMRPOQLATETITSSGFRLAMPPLTTADS	240	
QY	241	GYVVELVSAQPGARROQ	LPGNATDWIMAGLDPTDVALVPESNRLRPPQILRYR	300	
DB	241	GYVVELVSAQPGARROQ	LPGNATDWIMAGLDPTDVALVPESNRLRPPQILRYR	300	
QY	241	GYVVELVSAQPGARROQ	LPGNATDWIMAGLDPTDVALVPESNRLRPPQILRYR	300	
DB	241	GYVVELVSAQPGARROQ	LPGNATDWIMAGLDPTDVALVPESNRLRPPQILRYR	300	
QY	301	TRPEENGPRIRIYSHAR	PRSLRVSAPALGSAALGYHQFGPLRGEAQRVEVPAGRNC	360	
DB	301	TRPEENGPRIRIYSHAR	PRSLRVSAPALGSAALGYHQFGPLRGEAQRVEVPAGRNC	360	
QY	301	TRPEENGPRIRIYSHAR	PRSLRVSAPALGSAALGYHQFGPLRGEAQRVEVPAGRNC	360	
DB	301	TRPEENGPRIRIYSHAR	PRSLRVSAPALGSAALGYHQFGPLRGEAQRVEVPAGRNC	360	

```
QY 361 TTLOGIAGTAYLVVTAAFRSGRESALSAXACTPDGPPRPVPVPRAPTGTASREP 418
Db 361 TTLOGIAGTAYLVVTAAFRSGRESALSAXACTPDGPPRPVPVPRAPTGTASREP 418

RESULT 2
US-10-699-035A-20
; Sequence 20, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-20

Query Match 100.0%; Score 2154; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 2e-174;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPTALGLALSRLALASGAERGPASAPRGDMLFLDSSASVSHYEFSSVREFVQGL 60
Db 1 MLPTALGLALSRLALASGAERGPASAPRGDMLFLDSSASVSHYEFSSVREFVQGL 60

QY 61 VAPPLGTGALRASLVHVGSRPYTEFPFGQSSGGAADAVASAOQMGDTHTGLALVYA 120
Db 61 VAPPLGTGALRASLVHVGSRPYTEFPFGQSSGGAADAVASAOQMGDTHTGLALVYA 120

QY 121 KEQLFAEASGARPGVKLVWTTDGGSSDPVGPMPQELKDGLGVTFIVSTGRGNFLELSA 180
Db 121 KEQLFAEASGARPGVKLVWTTDGGSSDPVGPMPQELKDGLGVTFIVSTGRGNFLELSA 180

QY 181 AASAPAEKHLHFVDVDDHIIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLLTADS 240
Db 181 AASAPAEKHLHFVDVDDHIIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLLTADS 240

QY 241 GYYVIELVPSAQPGAAARQQLPGNATDWTIWAGLDPTDYDALVPESNVRLIRPOLLRVR 300
Db 241 GYYVIELVPSAQPGAAARQQLPGNATDWTIWAGLDPTDYDALVPESNVRLIRPOLLRVR 300

QY 301 TRPEAGGERIVTISHARRSLRVSWAPALGSAALGYHVOQPLRGGEAQRVEVPAGNVC 360
Db 301 TRPEAGGERIVTISHARRSLRVSWAPALGSAALGYHVOQPLRGGEAQRVEVPAGNVC 360

QY 361 TTLOGIAGTAYLVVTAAFRSGRESALSAXACTPDGPPRPVPVPRAPTGTASREP 418
Db 361 TTLOGIAGTAYLVVTAAFRSGRESALSAXACTPDGPPRPVPVPRAPTGTASREP 418

RESULT 3
US-10-699-035A-4
; Sequence 4, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT FILING DATE: 2003-10-31
; CURRENT APPLICATION NUMBER: US/10/699,035A
```

```
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-4

Query Match 76.8%; Score 1653.5; DB 4; Length 415;
Best Local Similarity 77.5%; Pred. No. 7.4e-132;
Matches 324; Conservative 32; Mismatches 59; Indels 3; Gaps 1;

QY 1 MLPTALGLALSRLALASGAERGPASAPRGDMLFLDSSASVSHYEFSSVREFVQGL 60
Db 1 MLPTALGLALSRLALASGAERGPASAPRGDMLFLDSSASVSHYEFSSVREFVQGL 60

QY 61 VAPPLGTGALRASLVHVGSRPYTEFPFGQSSGGAADAVASAOQMGDTHTGLALVYA 120
Db 61 VAPPLGTGALRASLVHVGSRPYTEFPFGQSSGGAADAVASAOQMGDTHTGLALVYA 120

QY 121 KEQLFAEASGARPGVKLVWTTDGGSSDPVGPMPQELKDGLGVTFIVSTGRGNFLELSA 180
Db 121 KEQLFAEASGARPGVKLVWTTDGGSSDPVGPMPQELKDGLGVTFIVSTGRGNFLELSA 180

QY 181 AASAPAEKHLHFVDVDDHIIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLLTADS 240
Db 181 AASAPAEKHLHFVDVDDHIIIVQELRGSILDMARPOQLHATEITSSGFRLLAMPPLLTADS 240

QY 241 GYYVIELVPSAQPGAAARQQLPGNATDWTIWAGLDPTDYDALVPESNVRLIRPOLLRVR 300
Db 241 GYYVIELVPSAQPGAAARQQLPGNATDWTIWAGLDPTDYDALVPESNVRLIRPOLLRVR 300

QY 301 TRPEAGGERIVTISHARRSLRVSWAPALGSAALGYHVOQPLRGGEAQRVEVPAGNVC 360
Db 301 TRPEAGGERIVTISHARRSLRVSWAPALGSAALGYHVOQPLRGGEAQRVEVPAGNVC 360

QY 361 TTLOGIAGTAYLVVTAAFRSGRESALSAXACTPDGPPRPVPVPRAPTGTASREP 418
Db 361 TTLOGIAGTAYLVVTAAFRSGRESALSAXACTPDGPPRPVPVPRAPTGTASREP 418

RESULT 4
US-10-699-035A-21
; Sequence 21, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-21

Query Match 76.8%; Score 1653.5; DB 4; Length 415;
Best Local Similarity 77.5%; Pred. No. 7.4e-132;
Matches 324; Conservative 32; Mismatches 59; Indels 3; Gaps 1;

QY 1 MLPTALGLALSRLALASGAERGPASAPRGDMLFLDSSASVSHYEFSSVREFVQGL 60
```

```
Db      1 MLFWTAFSVALSLRLALARSISIERGSTASDPQGDLLFLDSSASVSHYEFSSRVREFVGQL 60
QY      61 VAPLPCTGALRASLVHVSRRPYTEPPFGQSSGEAQAODAVRASAOQMGDTHTGLALVYA 120
Db      61 VMTMSGFPALRASLVHVSQPTTEFTFDYSSGQALRDLIRKAPRMDTHTGLALVYA 120
QY      121 KEOLFPAASGARPGVPKVLVWTDGSSDPVGPMPQELKDLGTVTIVSTGRGNFLELSA 180
Db      121 KEOLFPAASGARPGVPKVLVWTDGSSDPVGPMPQELKDLGTVTIVSTGRGNFLELSA 180
QY      181 AASAPAEKHLHFVDVDDLHIIVQELRGSITDAMQPQOLHASELSSGFRLSWPLLTADS 240
Db      181 AASAPAEKHLHFVDVDDLHIIVQELRGSITDAMQPQOLHASELSSGFRLSWPLLTADS 240
QY      241 GYYVLELVPSAOGAARROOLPQNATDWMAGIDPPTDVALVPSNRLLRPOILRYR 300
Db      241 GYYVLELVPSGKIATRRROOLPQNATSWTDDPDTDEVSLLPSSNHLARPQHVRYR 300
QY      301 TRPEAGPERIVISHARPSRLRVSMAPALGSAALGYHVOFGLRGGEAQREVPAGRNC 360
Db      301 TIQEEAGPERIVISHARPSRLRVSMAPALGSDALGYHVOFGLRGGEAQREVPAGRNC 360
QY      361 TTLQGLAPGTAYLVYTTAAFRSGRESALSAKACTPPGPPRPVRAPIPTGASREP 418
Db      361 TTVOGLTPTCTYLVYTTAAFRSGROALSAKACTASGARTR--AFQSMKPEAGPREP 415

RESULT 5
US-09-789-561-85
; Sequence 85, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-85

Query Match      50.3%; Score 1084; DB 3; Length 215;
Best Local Similarity 99.5%; Pred. No. 8.4e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 MLFWTALGLALSLRLALARSGERGPASAPRGDLFLDSSASVSHYEFSSRVREFVGQL 60
Db      1 MLFWTAXGLALSLRLALARSGARPPASAPRGDLFLDSSASVSHYEFSSRVREFVGQL 60
QY      61 VAPLPCTGALRASLVHVSRRPYTEPPFGQSSGEAQAODAVRASAOQMGDTHTGLALVYA 120
Db      61 VAPLPCTGALRASLVHVSRRPYTEPPFGQSSGEAQAODAVRASAOQMGDTHTGLALVYA 120
QY      121 KEOLFPAASGARPGVPKVLVWTDGSSDPVGPMPQELKDLGTVTIVSTGRGNFLELSA 180
Db      121 KEOLFPAASGARPGVPKVLVWTDGSSDPVGPMPQELKDLGTVTIVSTGRGNFLELSA 180
QY      181 AASAPAEKHLHFVDVDDLHIIVQELRGSITDAMRP 215
```

```
Db      181 AASAPAEKHLHFVDVDDLHIIVQELRGSITDAMRP 215

RESULT 6
US-09-833-245-2096
; Sequence 2096, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2096
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2096

Query Match      50.3%; Score 1084; DB 3; Length 215;
Best Local Similarity 99.5%; Pred. No. 8.4e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 MLFWTALGLALSLRLALARSGERGPASAPRGDLFLDSSASVSHYEFSSRVREFVGQL 60
Db      1 MLFWTAXGLALSLRLALARSGARPPASAPRGDLFLDSSASVSHYEFSSRVREFVGQL 60
QY      61 VAPLPCTGALRASLVHVSRRPYTEPPFGQSSGEAQAODAVRASAOQMGDTHTGLALVYA 120
Db      61 VAPLPCTGALRASLVHVSRRPYTEPPFGQSSGEAQAODAVRASAOQMGDTHTGLALVYA 120
QY      121 KEOLFPAASGARPGVPKVLVWTDGSSDPVGPMPQELKDLGTVTIVSTGRGNFLELSA 180
Db      121 KEOLFPAASGARPGVPKVLVWTDGSSDPVGPMPQELKDLGTVTIVSTGRGNFLELSA 180
QY      181 AASAPAEKHLHFVDVDDLHIIVQELRGSITDAMRP 215
Db      181 AASAPAEKHLHFVDVDDLHIIVQELRGSITDAMRP 215

RESULT 7
US-10-983-936-85
; Sequence 85, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/10/883,936
; PRIOR FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 85
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-85
```

```
Query Match          50.3%; Score 1084; DB 5; Length 215;
Best Local Similarity 99.5%; Pred. No. 8.4e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MLPWTALGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 60
DB 1 MLPWTAXGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 60
QY 61 VAPLPLGTGALPASLVHGSRRPYTEFPFGQSSGGAADAVASAQRMGDHTTGALVYA 120
DB 61 VAPLPLGTGALPASLVHGSRRPYTEFPFGQSSGGAADAVASAQRMGDHTTGALVYA 120
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDGLGVTIVISTGRGNFLELSA 180
DB 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDGLGVTIVISTGRGNFLELSA 180
QY 181 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 215
DB 181 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 215
```

```
RESULT 8
US-09-789-561-159
; Sequence 159, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043Pl
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 159
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-159
```

```
Query Match          50.3%; Score 1084; DB 3; Length 242;
Best Local Similarity 99.5%; Pred. No. 9.9e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MLPWTALGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 60
DB 28 MLPWTAXGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 87
```

```
QY 61 VAPLPLGTGALPASLVHGSRRPYTEFPFGQSSGGAADAVASAQRMGDHTTGALVYA 120
DB 88 VAPLPLGTGALPASLVHGSRRPYTEFPFGQSSGGAADAVASAQRMGDHTTGALVYA 147
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDGLGVTIVISTGRGNFLELSA 180
DB 148 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDGLGVTIVISTGRGNFLELSA 207
QY 181 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 215
DB 208 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 242
```

```
RESULT 9
US-10-883-936-159
; Sequence 159, Application US/10883936
; Publication No. US20050019866A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043Pl
; CURRENT APPLICATION NUMBER: US/10/883,936
; PRIOR FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/789,561
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 159
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-883-936-159
```

```
Query Match          50.3%; Score 1084; DB 5; Length 242;
Best Local Similarity 99.5%; Pred. No. 9.9e-84;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MLPWTALGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 60
DB 28 MLPWTAXGIALSLRLALARSAGRGPPASAPRGDLMFLDSSASVSHYEFSSRVREFVQGL 87
QY 61 VAPLPLGTGALPASLVHGSRRPYTEFPFGQSSGGAADAVASAQRMGDHTTGALVYA 120
DB 88 VAPLPLGTGALPASLVHGSRRPYTEFPFGQSSGGAADAVASAQRMGDHTTGALVYA 147
QY 121 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDGLGVTIVISTGRGNFLELSA 180
DB 148 KEQLFAEASGARPGVPKVLVWTTDGGSSDPVGPMPQELKDGLGVTIVISTGRGNFLELSA 207
QY 181 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 215
DB 208 AASAPAEKHLHFVDVDDLHIIVQELRGSILDAMP 242
```

```
RESULT 10
```



```
US-10-699-035A-2
; Sequence 2, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-699-035A-2
```

```
Query Match          42.4%; Score 913; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 2,3e-69;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      32 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
DB      1 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 60

QY      92 SSGEAAQADAVRASAOQMGDTHTGTLALVYAKQOLFPAASGARGVGVKVLVWTTDGGSSDPV 151
DB      61 SSGEAAQADAVRASAOQMGDTHTGTLALVYAKQOLFPAASGARGVGVKVLVWTTDGGSSDPV 120

QY      152 GPPMQLKDLGVTIVTSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211
DB      121 GPPMQLKDLGVTIVTSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
```

```
RESULT 14
US-10-699-035A-11
; Sequence 31, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: VA domain from WARP
US-10-699-035A-31
```

```
Query Match          35.8%; Score 772; DB 4; Length 181;
Best Local Similarity 83.4%; Pred. No. 2,3e-57;
Matches 151; Conservative 14; Mismatches 16; Indels 0; Gaps 0;
```

```
QY      32 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
DB      1 QGDLLFLDSSASVSHYEFSSRVREFVQGLVATWMSFGPALRASLVHVGSPHTTEFTFDQY 60

QY      92 SSGEAAQADAVRASAOQMGDTHTGTLALVYAKQOLFPAASGARGVGVKVLVWTTDGGSSDPV 151
```

```
DB      61 SSGQAIRDAIRVAPQRMGDTHTGTLALVYAKQOLFPEEAGARGVGVKVLVWTTDGGSSDPV 120
QY      152 GPPMQLKDLGVTIVTSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211
DB      121 GPPMQLKDLGVTIVTSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180

QY      212 A 212
DB      181 A 181
```

```
RESULT 15
US-10-699-035A-8
; Sequence 8, Application US/10699035A
; Publication No. US20040214349A1
; GENERAL INFORMATION:
; APPLICANT: Bateman, John
; APPLICANT: Fitzgerald, David
; TITLE OF INVENTION: A Molecular Marker
; FILE REFERENCE: A36056 PCT USA A 071838.0142
; CURRENT APPLICATION NUMBER: US/10/699,035A
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/AU02/00542
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: AU PR4701/01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-699-035A-8
```

```
Query Match          35.7%; Score 768; DB 4; Length 180;
Best Local Similarity 83.3%; Pred. No. 5e-57;
Matches 150; Conservative 14; Mismatches 16; Indels 0; Gaps 0;
```

```
QY      32 RGDLMFLDSSASVSHYEFSSRVREFVQGLVAPLPLGTGALRASLVHVGSRPYTEPPFGQH 91
DB      1 QGDLLFLDSSASVSHYEFSSRVREFVQGLVATWMSFGPALRASLVHVGSPHTTEFTFDQY 60

QY      92 SSGEAAQADAVRASAOQMGDTHTGTLALVYAKQOLFPAASGARGVGVKVLVWTTDGGSSDPV 151
DB      61 SSGQAIRDAIRVAPQRMGDTHTGTLALVYAKQOLFPEEAGARGVGVKVLVWTTDGGSSDPV 120

QY      152 GPPMQLKDLGVTIVTSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 211
DB      121 GPPMQLKDLGVTIVTSTGRGNFLELSAASAPAEKHLHFVDVDDLHIIVQELRGSLTD 180
```

```
Search completed: February 13, 2006, 13:12:40
Job time : 86.2776 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2006, 07:47:59 ; Search time 64.3077 Seconds
(without alignments)
537.392 Million cell updates/sec

Title: US-10-699-035a-6

Perfect score: 2154
Sequence: 1 MPTWALGIALSLRLAARS.....RPRPPVPRAPPGTASREP 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5/COMB.pep:*
2: /cgn2_6/prodata/1/1aa/6/COMB.pep:*
3: /cgn2_6/prodata/1/1aa/H/COMB.pep:*
4: /cgn2_6/prodata/1/1aa/PCTUS/COMB.pep:*
5: /cgn2_6/prodata/1/1aa/RE/COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384	17.8	77	2	US-09-513-999C-6277 Sequence 6277, App1
2	264	12.3	755	2	US-09-919-497-57 Sequence 57, App1
3	259	12.0	584	2	US-09-949-016-10340 Sequence 10340, A
4	259	12.0	584	2	US-09-949-016-10341 Sequence 10341, A
5	255	11.8	963	2	US-09-949-016-11519 Sequence 11519, A
6	255	11.8	963	2	US-09-949-016-11520 Sequence 11520, A
7	253.5	11.8	638	1	US-08-897-443-1 Sequence 1, App1
8	253.5	11.8	915	2	US-09-907-794A-34 Sequence 34, App1
9	253.5	11.8	915	2	US-09-905-125A-34 Sequence 34, App1
10	253.5	11.8	915	2	US-09-902-775A-34 Sequence 34, App1
11	253.5	11.8	915	2	US-09-906-700-34 Sequence 34, App1
12	253.5	11.8	915	2	US-09-903-603A-34 Sequence 34, App1
13	253.5	11.8	915	2	US-09-904-820A-34 Sequence 34, App1
14	253.5	11.8	915	2	US-09-909-064-34 Sequence 34, App1
15	253.5	11.8	915	2	US-09-905-381A-34 Sequence 34, App1
16	253.5	11.8	915	2	US-09-906-618-34 Sequence 34, App1
17	253.5	11.8	915	2	US-09-906-646-34 Sequence 34, App1
18	253.5	11.8	915	2	US-09-904-462-34 Sequence 34, App1
19	253.5	11.8	915	2	US-09-902-736A-34 Sequence 34, App1
20	253.5	11.8	915	2	US-09-906-722A-34 Sequence 34, App1
21	253.5	11.8	915	2	US-09-905-125A-34 Sequence 34, App1
22	245	11.4	956	1	US-08-897-443-3 Sequence 3, App1
23	237	11.0	452	2	US-09-914-259-34 Sequence 34, App1
24	233.5	10.8	496	1	US-08-463-180-37 Sequence 37, App1
25	233.5	10.8	496	1	US-08-001-078A-1 Sequence 1, App1
26	233.5	10.8	496	1	US-08-897-443-4 Sequence 4, App1
27	233.5	10.8	496	1	US-08-897-443-4 Sequence 4, App1

28	233.5	10.8	496	1	US-08-463-218-1 Sequence 1, App1
29	233.5	10.8	496	2	US-09-949-016-11306 Sequence 11306, A
30	233.5	10.8	496	4	PCT-US94-00253-1 Sequence 1, App1
31	232	10.8	405	2	US-09-312-283C-374 Sequence 374, App
32	228.5	10.6	481	2	US-09-914-259-36 Sequence 36, App1
33	223.5	10.4	1036	2	US-10-104-047-2812 Sequence 2812, App
34	221	10.3	486	2	US-09-914-259-35 Sequence 35, App1
35	221	10.3	486	2	US-09-976-594-278 Sequence 278, App
36	221	10.3	486	2	US-09-949-016-6216 Sequence 6216, App
37	221	10.3	507	2	US-09-949-016-9878 Sequence 9878, App
38	212.5	9.9	954	2	US-09-996-611D-1 Sequence 1, App1
39	202	9.4	1180	2	US-09-000-004A-2 Sequence 2, App1
40	198.5	9.2	171	2	US-09-996-611D-2 Sequence 2, App1
41	194	9.0	214	2	US-10-061-658-6 Sequence 6, App1
42	194	9.0	214	2	US-10-061-658-9 Sequence 9, App1
43	194	9.0	550	2	US-09-907-794A-227 Sequence 227, App
44	194	9.0	550	2	US-09-905-125A-227 Sequence 227, App
45	194	9.0	550	2	US-09-902-775A-227 Sequence 227, App

ALIGNMENTS

```
RESULT 1
US-09-513-999C-6277
; Sequence 6277, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumae Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59. US2, REG
; CURRENT APPLICATION NUMBER: US/09/513, 999C
; PRIOR APPLICATION NUMBER: 2000-02-24
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent. pm
; SEQ ID NO 6277
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 66
; OTHER INFORMATION: Xaa=Cys or Gly
US-09-513-999C-6277

Query Match          17.8% Score 384; DB 2; Length 77;
Best Local Similarity 98.7%; Pred. No. 1.7e-26;
Matches 76; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      108 MGDHTGIALVYAKQLFAEASGARPGVKLVWTTDGGSDPVGPMQELKDLGVTVPI 167
      1 MGDHTGIALVYAKQLFAEASGARPGVKLVWTTDGGSDPVGPMQELKDLGVTVPI 60
DB      61 VSTGRNPFLELSAASA 77
QY      168 VSTGRNPFLELSAASA 184
DB      61 VSTGRNPFLELSAASA 77

RESULT 2
US-09-919-497-57
; Sequence 57, Application US/09919497
; Patent No. 673883
; GENERAL INFORMATION:
; APPLICANT: Muller, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919, 497
; CURRENT FILING DATE: 2001-07-31
```

;; PRIOR APPLICATION NUMBER: US 60/221,735
;; PRIOR FILING DATE: 2000-07-31
;; NUMBER OF SEQ ID NOS: 100
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 57
;; LENGTH: 755
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-919-497-57

Query Match 12.3%; Score 264; DB 2; Length 755;
Best Local Similarity 29.5%; Pred. No. 1,5e-14;
Matches 62; Conservative 44; Mismatches 90; Indels 14; Gaps 3;

QY 30 APERGDLMLDSSASVSHYEFSEFVRBEFVGQVLAFLPLGTGALRASLVHVGSRPYTEPPFGHSS 88
DB 3 AAKADLVFVVDGSMISGIDENFNKIIISFLYSTGALNKIKITDGTQVAMVQFTDDPRTPEFL 62
QY 89 GQHSSEGAQAQDVRASAQMGDTHTGLALVYAKEQLFAEASGAPRGVPLVWVTDGSSD 148
DB 63 NAYKRETELDAIKHISYKGNKTKGKAIVYRDTFTAESGTRRGIPKVIIVITDGRSQ 122
QY 149 DPGVPMQELKDLGTVTFIVSTGRGNFLELSAASAPAEKHLHPVD-----VDDLHI 200
DB 123 DVNKRISREMOQDGYSTIFAIQVADADYSELVSTIGSKPSARHVPVDDPDPAFKKIEDELIT 182
QY 201 IVQELRGSLIDAMPQQLHATEITSSGFRLL 230
DB 183 FVCEHTASATCPV-----VHKDGLDLAGFCM 207

RESULT 3
US-09-949-016-10340
; Sequence 10340, Application US/09949016
; Patent No. 6812339

;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CLO01307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 10340
;; LENGTH: 584
;; TYPE: PRT
;; ORGANISM: Human
US-09-949-016-10340

Query Match 12.0%; Score 259; DB 2; Length 584;
Best Local Similarity 37.2%; Pred. No. 2.9e-14;
Matches 67; Conservative 25; Mismatches 84; Indels 4; Gaps 2;

QY 34 DLMFLDSSASVSHYEFSEFVRBEFVGQVLAFLPLGTGALRASLVHVGSRPYTEPPFGHSS 93
DB 348 DLVLVLDGSKSVRPQNFELVYKRFVNIQVDFLDVSPGRTVGLVQFSSVRTEFFPLGRYGT 407
QY 94 GEAADAVRASAQRMGDTHTGLALVYAKEQLFAEASGAPR---GVPKVLVWVTDGSSSDP 150
DB 408 AAEVQAQVLAVERMERTGTGLALRHMYEHSFSEAGAPRALNVPRGVLFTDGRSQDD 467
QY 151 VGPPEMQLKDLGTVTFIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGS 209
DB 468 ISVMAAKAKERGIWYAVGVGKAVAEALREIASPEALHVSVA PDGTMTHLENLRGS 527

RESULT 4
US-09-949-016-10341
; Sequence 10341, Application US/09949016
; Patent No. 6812339

;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CLO01307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 10341
;; LENGTH: 584
;; TYPE: PRT
;; ORGANISM: Human
US-09-949-016-10341

Query Match 12.0%; Score 259; DB 2; Length 584;
Best Local Similarity 37.2%; Pred. No. 2.9e-14;
Matches 67; Conservative 25; Mismatches 84; Indels 4; Gaps 2;

QY 34 DLMFLDSSASVSHYEFSEFVRBEFVGQVLAFLPLGTGALRASLVHVGSRPYTEPPFGHSS 93
DB 348 DLVLVLDGSKSVRPQNFELVYKRFVNIQVDFLDVSPGRTVGLVQFSSVRTEFFPLGRYGT 407
QY 94 GEAADAVRASAQRMGDTHTGLALVYAKEQLFAEASGAPR---GVPKVLVWVTDGSSSDP 150
DB 408 AAEVQAQVLAVERMERTGTGLALRHMYEHSFSEAGAPRALNVPRGVLFTDGRSQDD 467
QY 151 VGPPEMQLKDLGTVTFIVSTGRGNFLELSAASAPAEKHLHFV-DVDDLHIIVQELRGS 209
DB 468 ISVMAAKAKERGIWYAVGVGKAVAEALREIASPEALHVSVA PDGTMTHLENLRGS 527

RESULT 5
US-09-949-016-11519
; Sequence 11519, Application US/09949016
; Patent No. 6812339

;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CLO01307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 11519
;; LENGTH: 963
;; TYPE: PRT
;; ORGANISM: Human
US-09-949-016-11519

Query Match 11.8%; Score 255; DB 2; Length 963;
Best Local Similarity 27.9%; Pred. No. 1.3e-13;
Matches 68; Conservative 43; Mismatches 89; Indels 44; Gaps 4;

QY 5 TALGLALSURL-----ALAKSGARGPPASA----- 30

Db 32 TALPLLLALMKEMKLAGCFLILIGQIVLPAEAREBSRGRSISRGHARTHPOTALLSS 91
QY 31 ---PRGDMFLDSSASVSHYEFSSRVREPVGOVAPLPJGTGALRASLVHVGSRPYTEP 87
Db 92 CENKRADLVFIIDSSRSVTHDYAKVEFIVDILQFLDGPVTRVGLLQYGSTVKNES 151
QY 88 FCGHSSGEAAQODAVRASAOQMGDTHGTALVYAKEOLFPAASGARP---GVPEVLYWVWD 144
Db 152 LKTFKRKSEVERAVKMRHLSTGTMTGLAQYALNIAFSEBEGARPLRENVPRVIMIVTD 211
QY 145 GGSDDPVGPMPQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHL---HFVDDVLDHI 200
Db 212 GRPDQSVAAAKAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVLVANFSQIETLS 271
QY 201 IVOE 204
Db 272 VFQK 275

RESULT 6
US-09-949-016-11520
; Sequence 11520, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11520
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11520

Query Match 11.8%; Score 255; DB 2; Length 963;
Best Local Similarity 27.9%; Pred. No. 1,3e-13;
Matches 68; Conservative 43; Mismatches 89; Indels 44; Gaps 4;

QY 5 TALGALSLRL-----ALARSGAERGPPASA----- 30
Db 32 TALPLLLALMKEMKLAGCFLILIGQIVLPAEAREBSRGRSISRGHARTHPOTALLSS 91
QY 31 ---PRGDMFLDSSASVSHYEFSSRVREPVGOVAPLPJGTGALRASLVHVGSRPYTEP 87
Db 92 CENKRADLVFIIDSSRSVTHDYAKVEFIVDILQFLDGPVTRVGLLQYGSTVKNES 151
QY 88 FCGHSSGEAAQODAVRASAOQMGDTHGTALVYAKEOLFPAASGARP---GVPEVLYWVWD 144
Db 152 LKTFKRKSEVERAVKMRHLSTGTMTGLAQYALNIAFSEBEGARPLRENVPRVIMIVTD 211
QY 145 GGSDDPVGPMPQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHL---HFVDDVLDHI 200
Db 212 GRPDQSVAAAKAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVLVANFSQIETLS 271
QY 201 IVOE 204
Db 272 VFQK 275

RESULT 7
US-08-897-443-1
; Sequence 1, Application US/08897443
; Patent No. 5981263

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Kaser, Mathew
; TITLE OF INVENTION: HUMAN MATRILIN-3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,443
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: UTRSN0702
; CLONE: 681719
US-08-897-443-1

Query Match 11.8%; Score 253.5; DB 1; Length 638;
Best Local Similarity 31.6%; Pred. No. 9.9e-14;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERGPPASAPRGDMFLDSSASVSHYEFSSRVREPVGOVAPLPJGTGALRASL 75
Db 45 ALLESSCEN-----KRALVLFIIIDSSRSVTHDYAKVEFIVDILQFLDGPVTRVGL 98
QY 76 VHVGSRPYTEPFCGHSSEAAQODAVRASAOQMGDTHGTALVYAKEOLFPAASGARP-- 133
Db 99 LQYGSVKNESLKTFFKRSEVERAVKMRHLSTGTMTGLAQYALNIAFSEBEGARPLR 158
QY 134 -GVPEVLYWVWTDGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHL-- 190
Db 159 ENVPRIYIMIVTQGRPDQSVAAAKAKARDTGILIFAIQVGVDFNTLKSIGSEPHEDHVL 218
QY 191 --HFVDDVLDHIIVOE 204
Db 219 VANFSQIETLTSVFQK 234

RESULT 8
US-09-907-794A-34
; Sequence 34, Application US/09907794A
; Patent No. 6635468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,794A
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-907-794A-34

Query Match 11.8%; Score 253.5; DB 2; length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARGAERGPAPASPRGLMLLDSASVSHYERVRVEFGQLVAPLPLGTGLRSL 75

Db 45 ALLESSCEN-----KRADLVFTIDSSRVNTHDAKVEFLVDLQFLDGPVTRVGL 98
QY 76 VHVGSRPYTEFPFGQHSAGEAQAQDAVRASAGRMGDTHTGLALVYAKEOLFAEASGAP-- 133
Db 99 LQYSTVNEFSLKTFKRKSFEVERAVKMRHLSITGTMTGLAQVLMNTAFSABEGARPLR 158
QY 134 -GVPRVLVWVTDDGSSDPVGPMEQLKDLGTVTFIVSTGRGNFLELSAASAPAEKHL-- 190
Db 159 ENVPVRVIMIVTDGRQDSVAEVAARAKRDTGLIFAIQGVQDFNTLKSIGSEPHEDHVF 218
QY 191 --HFVDDHIIHVOE 204
Db 219 VANFSQIETLTSVFOK 234

RESULT 9
US-09-905-125A-34
Sequence 34, Application US/09905125A
Patent No. 6664376
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905,125A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30

```
/ PRIOR APPLICATION NUMBER: PCT/US99/28564
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/28565
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/30095
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: PCT/US99/30911
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US99/30999
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US00/00219
/ PRIOR FILING DATE: 2000-01-05
/ NUMBER OF SEQ ID NOS: 423
/ SEQ ID NO 34
/ LENGTH: 915
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-905-125A-34

Query Match      11.8%; Score 253.5; DB 2; Length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERPPASAPRGDLMFLDSSASVSHYEFKRVREFVQOLVAPLPGLTGALRASL 75
DB 45 ALLSSSCEN-----KRADLVFIIDSSRSVNTHDYAKVEFIVDIIQFLDIGDVTVRVGL 98
QY 76 VHVGSRPYTEFPFGQSSGEAAODAVASAOQRMGDTHTGLALVYAKEQLFAEASGARP-- 133
DB 99 LQYGSTVKNKEFSLKTRKSEVERAVKRMHLSGTMTGLAIQYALNIAFSEAEGRPLR 158
QY 134 -GVPKYLVWTDGSSDPVGPMPQELKDLGVTFVIVSTGRGNFLLSAASAPAEKRL-- 190
DB 159 ENVPRIYIMVTGDRPDQSVAEVAAKARDTGILIFALGVGVDPNTLKSIGSEPHEDHVF 218
QY 191 --HFVDVDDLHIIVQE 204
DB 219 VANFSQIETLTSVFOK 234

RESULT 10
US-09-902-775A-34
/ Sequence 34, Application US/09902775A
/ Patent No. 6686451
/ GENERAL INFORMATION:
/ APPLICANT: Genentech, Inc.
/ APPLICANT: Ashkenazi, Avi
/ APPLICANT: Botstein, David
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gerlitsen, Mary E.
/ APPLICANT: Goddard, A.
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Hillan, Kenneth, J.
/ APPLICANT: Kijavlin, Ivar J.
/ APPLICANT: Mather, Jennie P.
/ APPLICANT: Pan, James
/ APPLICANT: Peoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William, I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ FILE REFERENCE: 10466-14
```

```
/ CURRENT APPLICATION NUMBER: US/09/902,775A
/ CURRENT FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: PCT/US00/04414
/ PRIOR FILING DATE: 2000-02-22
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
/ PRIOR APPLICATION NUMBER: US 60/146,222
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: PCT/US99/20594
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ PRIOR APPLICATION NUMBER: PCT/US99/28313
/ PRIOR FILING DATE: 1999-11-30
/ PRIOR APPLICATION NUMBER: PCT/US99/28564
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/28565
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/30095
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: PCT/US99/30911
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US99/30999
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US00/00219
/ PRIOR FILING DATE: 2000-01-05
/ NUMBER OF SEQ ID NOS: 423
/ SEQ ID NO 34
/ LENGTH: 915
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-902-775A-34

Query Match      11.8%; Score 253.5; DB 2; Length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERPPASAPRGDLMFLDSSASVSHYEFKRVREFVQOLVAPLPGLTGALRASL 75
DB 45 ALLSSSCEN-----KRADLVFIIDSSRSVNTHDYAKVEFIVDIIQFLDIGDVTVRVGL 98
QY 76 VHVGSRPYTEFPFGQSSGEAAODAVASAOQRMGDTHTGLALVYAKEQLFAEASGARP-- 133
DB 99 LQYGSTVKNKEFSLKTRKSEVERAVKRMHLSGTMTGLAIQYALNIAFSEAEGRPLR 158
QY 134 -GVPKYLVWTDGSSDPVGPMPQELKDLGVTFVIVSTGRGNFLLSAASAPAEKRL-- 190
DB 159 ENVPRIYIMVTGDRPDQSVAEVAAKARDTGILIFALGVGVDPNTLKSIGSEPHEDHVF 218
QY 191 --HFVDVDDLHIIVQE 204
DB 219 VANFSQIETLTSVFOK 234

RESULT 11
US-09-906-700-34
/ Sequence 34, Application US/09906700
/ Patent No. 6723535
/ GENERAL INFORMATION:
/ APPLICANT: Genentech, Inc.
/ APPLICANT: Ashkenazi, Avi
/ APPLICANT: Botstein, David
/ APPLICANT: Desnoyers, Luc
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Db          45  ALLESCEN-----KRADLVITIDSSSNVTHDYAKKEIVUILLQFLDIGPVTYVGL 98
Oy          76  VHVGSBPYTEPPFGHSGSEAAQDAVRASAPRMGDTHTGLALVYAKSQLFNASGARP-- 133
Db          99  LQSGSTVKNKEFSLKTPFKRSEVERAVKEMRHLSTGTMTGLAIQYALNINAFSEAGARPLR 158
Oy          134 -GPKVLWMTYDGGSSDPFGPMQELKDLGTVFIVSTVGRGNFLELSAASAPAEKLL-- 190
Db          159 ENNPRVIMVTYDGRPDSDVAEVAARABDTGLIFAIAGVDVFTLMSIGSEPHEDHVFL 218
Oy          191 --HFVDVDDLHIVQE 204
Db          219 VANFSQLETLTISYFQK 234

RESULT 12
US-09-903-603A-34
Sequence 34, Application US/09903603A
Patent No. 6767995
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: GNE.1618P2C12
CURRENT APPLICATION NUMBER: US/09/903,603A
PRIOR FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564

```

```

CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-904-920A-34

Query Match      11.8%; Score 253.5; DB 2; Length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY      16 ALARSGAERPPASAPRGDLMFLIDSSASVSHYEFSSVREFVGQVAPRLPGTALRLAST 75
Db      45 ALLESSCEN-----KRADLVFIIDSSRSVNTHDYAKVEFIYDILFLDGPVTVTGL 98
QY      76 VHGGSRYETFPFGQHSGEADQAVASAKRMGDHTGLALVYAKQQLFAEASGAPR- 133
Db      99 LQIGSTYAKNEFSLKTFGRKSEVEAVGRMHLSTGTGTGLIQYALINIAFSEAGAPLR 158
QY      134 -GVPKVLVWTTDGGSSDPVGPMPQELKDLGVTVFIVSTGNGFTELSSAAASAPAEKHL- 190
Db      159 ENVPKRVIMTVTDGRQDQSVAEVAAKADDTGLIFALIGVGVDFTLKSISSEPHEDHVFLL 218
QY      191 --HFVDVDDLHIYQE 204
Db      219 VANFSQIETLTSVFOK 234

RESULT 14
US-09-909-064-34
Sequence 34, Application US/0909064
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Denoyers, Luc
APPLICANT: Eaton, Dan L.

```

APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltzen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Goddard, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/909,064
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 34
LENGTH: 915
TYPE: PRT
ORGANISM: Homo sapiens
US-09-909-064-34
Query Match 11.8%; Score 253.5; DB 2; Length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;
QY 16 ALARSAERPPAPRGDMLPILDSASVSHSEFVREFOQVAPLEIGGALRSL 75
DB 45 ALLESSCEN-----KRAADVFLIDSSRSVNTHDYAKVKEIVDILQFLDIDGPDVTRVGL 98

QY 76 VHWGSRPYTERPPFGCHSGGEAAQDAVRASAOQMDTHTGLALVYAKEQLFAEASGARP-- 133
DB 99 LQYGSTVAKNEFSLKTFKRKSEVERAVKRMRLSTGTMTGLAIQYALNTAFSEAEGRPLR 158
QY 134 -GVPRVLVWMTDGGSSDPVGPMPQELKDLGVTFIVSTGRGNPLLSAAAPAKPKHL-- 190
DB 159 ENVPFVIMTVTDGRQDSVAEVAAMAKADPTGILITPAIGVGQYDFNTLKSIGSEPHEDHVF 218
QY 191 --HFVDVDDLHIIVQE 204
DB 219 VANFSQIEFTLTSVFX 234
RESULT 15
US-09-905-381A-34
Sequence 34, Application US/09905381A
Patent No. 6818746
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltzen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905,381A
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 34
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-381A-34

Query Match 11.8%; Score 253.5; DB 2; Length 915;
Best Local Similarity 31.6%; Pred. No. 1.6e-13;
Matches 62; Conservative 36; Mismatches 85; Indels 13; Gaps 3;

QY 16 ALARSGAERGPAPASPRGDMFLDSSASVSHYEFSPREFFVGQLVAPLPLGTGALRASL 75
Db 45 ALLESSCEN-----KRALVFIIDSRSVNTHDYAKVEFIYDILQFLDIGPDVTRVGL 98
QY 76 VHVGSRRPYTEFPFGQHSSEGAQDAVARASAORMGDTHTGLALVYAKEQLFAEASGARP-- 133
Db 99 LQYGSTVKNEFFSLKTFPKRSEVERAVRMRHLSTGTWTGLAIQYALNIATFSEAEAGARPLR 158
QY 134 -GVPKVLVWVTTDGGSSDPVGPMPQELKDLGVTVFIVSTGRGNFLELSAASAPAEKHL-- 190
Db 159 ENVPRIYIMIVTDRPDQSVAEVAAKARDTGILLFAIGVGQVDENFTLKSIGSEPHEDHVL 218
QY 191 --HFVDVDDLHIIVOE 204
Db 219 VANFSQIETITLTSVFQK 234

Search completed: February 13, 2006, 07:50:52
Job time : 65.3077 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 13, 2006, 13:39:36 ; Search time 9.24221 Seconds
(without alignments)
2243.517 Million cell updates/sec

Title: US-10-699-035A-5

Perfect score: 2380
Sequence: 1 atgtccctccgacgagcgct.....ccgcacccgctgacgcgttaa 1254

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-G/abses/ABSGWB.spool/US10699035/runat 13022006 062447 25513/app.query.fasta_1
-DB=Issued_Patents_AA -QPM=fasten -SUFFIX=rai -TINMATCH=0.1 -LOOEXT=0
-LOOEXT=0 -ONITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs02p -USER=US10699035 @CGN 1.1 101 @runat 13022006 062447 25513
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPLLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FCAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*

- 1: /cgnt_6/prodata/1/iaa/5/COMB.pep.*
- 2: /cgnt_6/prodata/1/iaa/6/COMB.pep.*
- 3: /cgnt_6/prodata/1/iaa/H.COMB.pep.*
- 4: /cgnt_6/prodata/1/iaa/PCUS.COMB.pep.*
- 5: /cgnt_6/prodata/1/iaa/RE.COMB.pep.*
- 6: /cgnt_6/prodata/1/iaa/backfilest1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384	16.1	77	2	US-09-513-999C-6277, Ap
2	287.5	12.1	860	2	US-09-252-991A-28607, A
3	281.5	11.8	663	2	US-09-252-991A-10843, A
4	279	11.7	511	2	US-09-252-991A-26078, A
5	279	11.7	2294	2	US-09-252-991A-117231, A
6	275.5	11.6	917	2	US-09-252-991A-203101, A
7	272	11.4	650	2	US-09-252-991A-20375, A
8	270	11.3	492	2	US-09-252-991A-28339, A
9	269.5	11.3	689	2	US-09-252-991A-31790, A
10	269	11.3	1706	2	US-09-252-991A-31760, A
11	268	11.3	653	2	US-09-252-991A-30843, A
12	267	11.2	638	2	US-09-252-991A-27068, A

13	266	11.2	755	2	US-09-919-497-57, A	Sequence 57, Appl
14	264.5	11.1	783	2	US-09-252-991A-18035, A	Sequence 18035, A
15	263.5	11.1	590	2	US-09-252-991A-19127, A	Sequence 19127, A
16	261	11.0	369	2	US-09-252-991A-25394, A	Sequence 25394, A
17	259.5	10.9	642	2	US-09-252-991A-24767, A	Sequence 24767, A
18	259	10.9	584	2	US-09-949-016-10340, A	Sequence 10340, A
19	259	10.9	584	2	US-09-949-016-10341, A	Sequence 10341, A
20	259	10.9	819	2	US-09-252-991A-19569, A	Sequence 19569, A
21	258	10.8	808	2	US-09-252-991A-32826, A	Sequence 32826, A
22	257	10.8	1228	2	US-09-252-991A-17764, A	Sequence 17764, A
23	256	10.8	2294	2	US-09-252-991A-17231, A	Sequence 17231, A
24	255.5	10.7	302	2	US-09-252-991A-19529, A	Sequence 19529, A
25	255	10.7	963	2	US-09-949-016-11519, A	Sequence 11519, A
26	255	10.7	963	2	US-09-949-016-11520, A	Sequence 11520, A
27	253.5	10.7	638	1	US-08-897-443-1, A	Sequence 1, Appl1
28	253.5	10.7	770	2	US-09-252-991A-30323, A	Sequence 30323, A
29	253.5	10.7	915	2	US-09-907-794A-34, A	Sequence 34, Appl
30	253.5	10.7	915	2	US-09-905-125A-34, A	Sequence 34, Appl
31	253.5	10.7	915	2	US-09-902-775A-34, A	Sequence 34, Appl
32	253.5	10.7	915	2	US-09-906-700-34, A	Sequence 34, Appl
33	253.5	10.7	915	2	US-09-903-603A-34, A	Sequence 34, Appl
34	253.5	10.7	915	2	US-09-904-920A-34, A	Sequence 34, Appl
35	253.5	10.7	915	2	US-09-909-064-34, A	Sequence 34, Appl
36	253.5	10.7	915	2	US-09-905-381A-34, A	Sequence 34, Appl
37	253.5	10.7	915	2	US-09-906-618-34, A	Sequence 34, Appl
38	253.5	10.7	915	2	US-09-906-646-34, A	Sequence 34, Appl
39	253.5	10.7	915	2	US-09-904-462-34, A	Sequence 34, Appl
40	253.5	10.7	915	2	US-09-902-736A-34, A	Sequence 34, Appl
41	253.5	10.7	915	2	US-09-906-722A-34, A	Sequence 34, Appl
42	253.5	10.7	956	2	US-09-949-016-6215, A	Sequence 6215, Ap
43	253	10.6	405	2	US-09-312-283C-374, A	Sequence 374, App
44	252.5	10.6	681	2	US-09-252-991A-24567, A	Sequence 24567, A
45	252.5	10.6	1065	1	US-08-642-255-72, A	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-09-513-999C-6277, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Maline Edwards, J.B.
; APPLICANT: Duclet, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513, 999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; NUMBER OF SEQ ID NOS: 1999-02-26
; SOFTWARE: Patent.pm
; SEQ ID NO 6277
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 66
; OTHER INFORMATION: Xaa=Cys or Gly
US-09-513-999C-6277
Alignment Scores: 1,616-18
Pred. No.: 384.00
Score: 98.7%
Percent Similarity: 98.7%
Best Local Similarity: 16.1%
Query Match: 2
DB: 0
US-10-699-035A-5 (1-1254) x US-09-513-999C-6277 (1-77)


```

RESULT 6
US-09-252-991A-25101
; Sequence 25101, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25101
; LENGTH: 917
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25101

```

```

Alignment Scores:
Pred. No.: 6,06e-11 Length: 917
Score: 275.50 Matches: 165
Percent Similarity: 31.9% Conservative: 21
Best Local Similarity: 28.3% Mismatches: 198
Query Match: 11.6% Indels: 199
DB: 2 Gaps: 31

```

US-10-699-035A-5 (1-1254) x US-09-252-991A-25101 (1-917)

```

QY 3 GCTCCCTGAGAGCGGCTCGGCT-----GGCCCTGAGCTTCGCGCTGCGCTGCGCG 56
DB 362 AIAATTCGyAAGlGlyProSerProAlaArgGlyGlyArgGlyAlaAAGlGlyGlyAspGly 381
QY 57 GAGCGCGCGAGCGGCTGACACGATCAAGCCCCCGAGGAGGACCTGATGTTCTGCT 116
DB 382 AATGAGlAGlAArgGlyAthrAsp-----GlyGlyProGluAlaProAla 397
QY 117 GGACAGCTCAGCCAGCT-----CTCTCAGCA 143
DB 398 ProAGlAArgProArgAspValGlyGlyLeuArgProAlaAlaThrLeuArgAla 417
QY 144 CGAGTTCTCCGGGTTGCG-----GGCTCCACTGCCCCCTGG 200
DB 418 ArgArgArgProGlyGlyArgArgGlyAspProGlnSerArgProAlaArgGlyAspPro 437
QY 162 -----GGAGTTGTGGGAGCTGCT--GGCTCCACTGCCCCCTGG 200
DB 438 ArgArgCyAArgArgProGlyArgAlaProAlaThrGlySerGlyArgArgAlaProGly 457
QY 201 CACCGGGGCGCTCGGCGCAAGTCGTGTCAGCTGGGCGGACATACCGGAGTTCTCC 260
DB 458 ArgArgSer-----GlyGlnGlyAlaProGlyGlnGlyGlyGlyPheArgArgPro 474
QY 261 CTT-----CGGCGACGACAGCTCGGCTGAGGCTGCGCCAGATGCGCT 302
DB 475 AlaValArgAspArgGlnHisArgProAlaArgHisProArgGlnHisAlaAspGly 494
QY 303 GCGTCTTCTGCCACGCGCATGGGTGACACCACTGCGCTGGCGCTGCTGATGCCA 362
DB 495 AlaLeuArgHisProAlaAspArg--ArgGlnHisPhePro-----507
QY 363 GGAACAGCTGTTCTGGAAGCATCAGATGCCCGGCGAGGGGTGCCAAGTGTGCT-- 419
DB 508 -----LeuGlnArgValProThrGlnGlyArgProGlyAlaGlyGly 521
QY 420 -----GTGGTGACAGATGGCGGCTCCAG 443
DB 522 AlaAspArgProValSerAlaGlnHisArgTyrProLeuSerThrAspArgProGlyLeuAla 541

```

```

QY 444 CGACCTCT-----GGGCCCCCATGTCAGAGCTCAAGA-----479
DB 542 ArgArgCyAArgGlyAspProGlyAlaProAlaAlaAlaArgAlaGlyAlaArg 561
QY 480 -----CGGGGCGTCACCGCTTTCATTGTCAGCAACGGCGAGGACATCTTCT 527
DB 562 ArgLeuAlaProProGlyArgAlaGlyGlyAspGlyGlnSerArgGlySerProAlaPro 581
QY 528 GGA---GCTGTACGCGCTGCTGAGCCCTGCGCGAGAGGACCTGCACTTGTGAGCT 584
DB 582 GlyArgArgAlaSerArgProAspGlnProAlaAlaGlyLeuPheProPheLeuValGlyAla 601
QY 585 GGATACCTGCACATCATTTG-----605
DB 602 -----ThrAlaGlyArgCyAAspSerLeuArgArgGlnArgLeuAlaTyrGlnLeuVal 619
QY 606 -----CSAAGCTGAGGGGCTGATCTGCGATGCGGCGCGAGAGCTTCA 653
DB 620 ArgProArgHisProHisAlaProGly--HisAlaArgLeuAlaValGlyGlnAla---637
QY 654 TGCCACGAGATCAAGTCCAGCG-----677
DB 638 -----GlyAspHisGlyGlnArgArgAlaLeuArgHisArgArgGlnAlaArgLeuSer 655
QY 677 -----677
DB 656 AlaAlaAlaGlyGlyAspGlyArgArgArgArgHisAlaAspGlyArgGlnArg 675
QY 678 -----CTTCGCGCTGCGCTGCGC-----ACCCCT 701
DB 676 ThrGlnHisArgGlyAlaAlaValLeuAlaLeuGlyLeuAlaSerLeuHisArgAlaGly 695
QY 702 GCTGACCGGAGA-----CTCGGCTGATGATGCTGGA-----GCT 737
DB 696 AlaGlnGlnArgArgSerGlnProGlyAspLeuGlyAlaAlaCyAAspProGlyArgArgPro 715
QY 738 GGTGCCGAGCGCGCGCGGCGGCTGCAAGACGCGAGAGCTGCGCGAGGAAAGCGCAGGA 797
DB 716 GlyValGlnProGlyAlaGlyGlySerAspArgPheProLeuArgAlaLeuCyAAspAla 735
QY 798 CTGATCTGGGCGCGGCTCGAACCC-----GGAACAGGA 830
DB 736 ArgLeuGlnGlnHisArgArgGlySerProGlyGlnHisArgArgGlyArgGly 755
QY 831 CTACGACGT-----GGCGCTAGTGTGATGTCACAGCTGCGCTGAGGCGCCA 881
DB 756 LeuArgArgArgProProGlyAlaAlaGlyGlyGlyGlnHisArg-----ProGlnArgAla 773
QY 882 GATCTGCGGGGTGCGCGCGCGCGAGAGGCGCGGCGCAAGGACATGTCATCTCCCA 941
DB 774 AlaAlaAlaAlaAlaHisGlnPheArgAlaGlyGlnAlaProAlaThrArgHisAlaAla 793
QY 942 ---GGCGCGCGCGCG-----CAGCTCCGCTGAGTTGGGCGCGCGCGCTGG 986
DB 794 GlyArgSerGlnAlaLeuGlyGlyAspSerProGlnArgProAlaGlyPheArgAlaAla 813
QY 987 CTACGCGCGCGC---GCTCGCTACACAGTGCAGTTGCGGCGCGCTGCGGCGGAGAGC 1043
DB 814 LeuSerLeuValAlaValAlaAlaLeuSerArgArgGlyArgGlyAlaThrGlyArg--Pro 832
QY 1044 GACGCGGTGAGAGGTGCGCGCGCGCGCGAGACCTGCACACAGCTGCGAGGCGCTGCGCGGG 1103
DB 833 ValArgGlyGlyGlyArgSerAlaProAlaAlaArgArgArgSerGlyAlaGlyGlnAsp 852
QY 1104 CACGCGCTACCTGAGTGCAGTGCAGCGCGCTT---CCGCTCGGCGCGCGAGAGCGCCT 1160
DB 853 ArgArgSerProGlyArgPheProArgSerValArgProAspGlyProAlaGlyGlySer 872
QY 1161 GTCCGCAAGGCTTGCAGCGCGA-----1184
DB 873 AlaArgArgGlyAlaAspArgAlaArgArgSerArgArgGlyArgArgProAlaValArgSer 892
QY 1185 -----CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGAC 1235

```

Db 893 GYAlAAlaArgAlaAlaProArgProPheProArgProGlyAlaArgAlaThrAspGly---- 911
QY 1236 CGGCAGCCG 1244
Db 912 ArgThrPro 914

RESULT 7
US-09-252-991A-20375
; Sequence 20375, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20375
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20375

Alignment Scores:
Pred. No.: 9,84e-11 Length: 650
Score: 272.00 Matches: 165
Percent Similarity: 32.9% Conservative: 22
Best Local Similarity: 29.0% Mismatches: 186
Query Match: 11.4% Indels: 196
DB: 2 Gaps: 33

US-10-699-035A-5 (1-1254) x US-09-252-991A-20375 (1-650)
QY 21 CGGCTGAGCCCTGAGCTTGGCGGTGGCGGCGGAGCGCGGAGCGCGGATCCACC 80
Db 6 ArgProGlyPro-----AlaGlyProAlaGlyGlyGlnArgArgArgAlaAspHisLeu 23
QY 81 AGCATCAGGCCCCCGAGGGAGCCTGATGTTCTGCTGAGCAAGCTCAGCCAG----- 131
Db 24 ArgValCysArgArgArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPro 43
QY 132 -----CGTCTCTACCTACGAGTTCTCCCGGGTTCCGGAAGTTGTGGGCA 176
Db 44 GlyProArgProLeuArgProLeuArgProArgValLeuPheGlnArgArgGlnArgGly--- 62
QY 177 GCGGTGGGCTCCACCTGCGGCGGACCGGGGCGCTGCGGAGCTGGTGCAGTGGG 236
Db 63 -----GlyProGlnHisArg---ProGlnArgGlnProGlyAla---Gly 75
QY 237 CAGTCGGCCATACCCGAGTTCCTTCGCGCAGCAGAG----- 275
Db 76 GlnAsp-----LeuArgProArgGlnArgArgGlyArgProGlyVal 89
QY 276 CTCGGGTGAGGCTGCCAGGA-----TCGGTGGC 305
Db 90 AlaGlyAlaGlyGlnProAlaGlyArgGlnLeuValGlnTrpGlnArgCysGlyAla 109
QY 306 TGGTTCGCGCAGCGAGTGGGTGACACCCACACTGGCGGCGGTGCTATGCCAAGA 365
Db 110 Ala---AlaProAlaGlnArgArgArgProArgGlnProArgProGlyArgLeuGlnPro 128
QY 366 ACAGCTGTGTTCTGAAGCATCAGGTGC-----CCGGCCAGGGGTGCCAAGT 413
Db 129 ArgHisAspLeuGlyGlnArgArgCysAlaAspProArgProGlnProGlyAlaGlnGly 148
QY 414 GCTGGTGTGGGTACAGATGGCGGCTCCAGCCACCTGTGGGCCCCCCTGAGAGAGCT 473

Db 149 ArgHisProArgArgArgGlnArgSerGlnArgArgLeuArgProGlyGlyAlaMetAla 168
QY 474 CAGGACCTGGCGCTCAC-----CGTGT 497
Db 169 GlyLeuProAlaArgHisLeuArgGlnProArgProAlaTrpArgTrpArgAlaArgArg 188
QY 498 CATTTGCAGCACCGGCGGAGGCACTTCCTGAGAGCTGTACGGCGCTCAGCCCCCTTC 557
Db 189 ArgArgGlyGlyArgProAlaThrAspProArgGlyAlaGlyArgArgArgHisSer 208
QY 558 CGAAGAGCAGCTGACCTTTGTGACGATGACCTGACATCATTTGTCCA----- 608
Db 209 AlaGlyGlnProAlaGlyArgHisArgGlyArgGlyGlyArgArgLeuProGlnArgGln 228
QY 609 AGAGCTGAGGGGCTCCATTTCTGCGATGCGGGCGCAGCAGCTCCATGCCAGCAGAT--- 665
Db 229 ArgAlaGlnGlyLeu-----LeuArgAlaAlaGlyArgCysHisGlyGlyArg 244
QY 666 -----CACGTC-----CAGCGGCTTCG 683
Db 245 ArgProGlyAspLeuArgArgHisLeuArgLeuArgArgArgGlnArgGlnArgGln 264
QY 684 CCGGGCTGGCCACCGCTGCTGACCGC-----AGACTC 716
Db 265 AlaArgArgProArgProGlyAspArgProGlyProArgValGlnArgProHisSerArg 284
QY 717 GGGCTACTATGTGCTGGA-----GCTGGTGC 743
Db 285 GlyArgAlaAlaAlaGlyGlnArgProGlyGlnArgArgGlyValAlaAspAlaGlyArg 304
QY 744 CAGCGCCCGCGGGGGGCTGCAGAGCGCAGCAGCTGCCAGGGAAGCCAC----- 794
Db 305 HisGlnProAlaAlaGlyGlnArgArgCysAlaGlyAlaLeuArgArgHisAlaArgAla 324
QY 795 -----GGACTGGAATCTGGCGGCGCTGACCGGAGAC 827
Db 325 AlaProProAspAlaAspArgGlnArgGlyLeu-----GlnProArgPro----- 339
QY 828 GGACTACGACGTGGCGCTAGTGCCTGATGCCAAGTCCAGCTGCCCTCTGAGGCCCGCATCT 887
Db 340 ---LeuArgArgGlyProAla-----ValLeuArgArgProProAspSerAspArgGln 356
QY 888 GCGGCTGCGCAGCGGCGCAGAGAGGCGGGCGCAGAGCCATCTCATCTCCACGCCCG 947
Db 357 ArgGlyGly---SerArgArgThrGlyGlnGlnHisAlaGlyAlaGlnArgProArgPro 375
QY 948 -----GCCGCGCAGCGCTCCGCGGAGTGGGCGCCCGACCGCT 983
Db 376 AspSerProArgArgArgProGlnArgAlaGlnProArgProAspProLeuProGlyThr 395
QY 984 GGGCTCAGCGCG----- 995
Db 396 GlyThrGlnArgLeuGlnGlnProAlaArgArgArgProTyGlnArgHisArgArgArg 415
QY 996 -----GGCGCTCGGCTACCA-----CGTGCA 1016
Db 416 AlaArgLeuLeuGlyArgGlnLeuProTyProAlaValAlaAspAlaAspLeuArgGly 435
QY 1017 GTTCGGGCGCTGCGGGGCGGGGAGCGCA-----CGCGGTGGA----- 1055
Db 436 SerGlnSerAlaSerGlyProAlaGlyGlyArgAlaProAlaGlyGlnLeuGln 455
QY 1056 -----GGTGCCCGCGGCGCGCACTG----- 1076
Db 456 ArgAlaGlnLeuLeuGlnArgArgTrpGlnGlnGlyArgArgPheProHisLeuAlaArgSer 475
QY 1077 CACCAAGCTGCAGGG---CTGGCGCC----- 1100
Db 476 GlnHisGlyArgGlyValProThrAlaGlnSerGlnAspArgArgGlyAspProArgLeu 495
QY 1101 -----GGGCAAGGCTACT---GGTGACGTGACCGCGGCTTCGCGGCGCGCA 1151
Db 496 GlnGlyGlyHisArgArgProAspGlyAspArgGlnArgArgLeuArgArgValGlnArg 515

```

Oy      1152 GAGCGCGCTGTCCGCAAGCCTGCAC---GCCGACGGCCCGCGCCGCCACGCC 1208
           |||          |||          |||          |||          |||
Db      516 HisArg-----ArgProGlyGlnArgProGlnArgGlnProTrnProGlyProAlaLeu 533
                                   :::::|||||||:::
Oy      1209 CGTCCCCCGCGCCCGCACCCCGGGGAC 1235
           :::::::|||||||:::
Db      534 CysLeuArgGlnProGluProGlyGlu 542

RESULT 8
US-09-252-991A-28339
; Sequence 28339, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28339
; LENGTH: 492
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28339
```

```

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31790
; LENGTH: 689
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31790

Alignment Scores:
Pred. No.: 1,47e-10 Length: 689
Score: 269.50 Matches: 159
Percent Similarity: 31.5% Conservative: 14
Best Local Similarity: 28.9% Mismatches: 158
Query Match: 11.3% Indels: 219
DB: Gaps: 33

US-10-699-035A-5 (1-1254) x US-09-252-991A-31790 (1-689)
QY 12 GACGGCGCTCGG-----CCGGCCCTGAGCTTGGCGCTGGCGCTGGCGCG 56
DB 194 AaPGLYAlaArgAlaAlaLeuArgAlaAaPGLYArgArgLeuSerProArgAlaArgLeu 213
QY 57 GACGGCGCGGAGCGGCTCCACAGCATCAGCCCGCGAGGGA----- 101
DB 214 ArgArgGlnGlyHisArgSerProProAlaArgLeuProAlaGlyArgThrGlnAlaGlu 233
QY 102 CCTGATGTTCTT---CTGGACAGCTCAGCCGCTCTCTACACAGATTCTCCGGGT 158
DB 234 ProAaPLeuProGlyAlaAaPHisHisArgGlnArg----- 245
QY 159 TCGGGAGTTTGGGGGCGACT-----GGTGGCTCCAGTCCCTGGGACCGGGGCTT 212
DB 246 GYAlaValAlaGlyAaPHisHisArgArgGlyAlaGlyArgAlaProGlyArgProGlyLeu 265
QY 213 GCGTGCAGCTGTGTGCACGTGGGCGAGTCCGCGCATACACCGATTCCCTCGCGCAGCA 272
DB 266 GYAlaArgLeuGly----- 270
QY 273 CAGCTCGGGTGAAGCTGCCCGAGATCGGCTGCTTCTGCCCGACGATGGGTGACAC 332
DB 271 GlnLeuAlaArgGlyAaPProArg-----ProCysHisGly--His 283
QY 333 CGACAC----- 338
DB 284 ProHisGlyAaPGLYAlaGlyArgProAlaValLeuGlnGlyArgArgHisHisArgProAaP 303
QY 339 -----TGG---CTGGCGCTGGTCTATGCCAA-----GAAACAGCTGTT 374
DB 304 ArgArgTrpLeuProArgArgArgGlyLeuHisGlnProLeuArgArgAlaGlyAlaProVal 323
QY 375 TCGTGAAGATCAGTGGTCCCGCGAGGGGTGCCAAAGTCTGTGTGGGTGACAGATGG 434
DB 324 GlnArgArgGlyArgArgGlyAlaArgAlaGlnGlnGlyProGlyGlyProAlaArgAla 343
QY 435 CGGCTCCAGCAGACCTGTGGGCCCCCGCATTCAGAGAGCTCAAGAGACTGGGCTGACCGT 494
DB 344 AlaLeuArgAaPProArgArgProProHisAlaAlaLeu-----GlyGlnHisArg 360
QY 495 GTTCATTGTCAAGACCGCGGAGCAACTTCTGAGCTGTGACCGCTGCTCAGCCCC 554
DB 361 ProAlaArgArgCysArgProArg-----ProGlyAlaTrpArgArg----- 374
QY 555 TGC CGAAGAAGACCTGCACTTGTGAGCTGATGACTGACATCAT----- 602
```

```

DB 375 -----GlyArgGly--ProValProHisArgSerAlaVal 385
QY 603 -----TGTCCAGAGCTGAGGGGCTTCATTCT-----CGCATGG 638
DB 386 HisAaPArgArgProLeuProGlnArgGlnGlyThrAlaGlyAaPLeuProArgAlaHis 405
QY 639 GCCGAGCAGCTCCATGCCACGAGATCAC----- 668
DB 406 GlnCysLeuProProAlaAlaGlyAaPHisHisArgProGlyTyArgArgArgGlnGly 425
QY 669 -----GTCCAGCGGCTT----- 680
DB 426 AlaValLeuLeuProAaPGLYAlaGlnProValProAlaGlnAlaArgHisProArg 445
QY 681 -----CCGCTGAGC---CTGGCACCCCTGTGACCGGACAGCTCGGCTACTATGT 728
DB 446 HisProArgProProGlyAaPLeuProGlyProAaPProArgHis----- 460
QY 729 GCTGAGCTGTGCCAGCGGCCAGCCGAGGAGCTGCMAAGACCGCAGACTGCCAGGAA 788
DB 461 -----AlaGlnGlyGlnArgArgThrGlyGlnProAlaTyProAlaAlaAaP----- 476
QY 789 CGCCACGAGACTGATCTGGGCGCGCTGACCCCGACACGAGCTAGAGCTGGCGCTAGT 848
DB 477 -----AaPLeuArgHisProArgAlaGlyArgGly----- 486
QY 849 GCCTGATGTCACAGTGGCGCTCTGTGAGCCCGCAGATCTCT-----GCGGATGG 896
DB 487 -----ProAlaProAaPProProArgLeuGlnGlyArgGlyAla 498
QY 897 CACGCGGCCAGAGAGCGCGGCGCAGAGCGCATGCTCAT----- 935
DB 499 ArgArgGlyArgGlyHisHisArgHisAlaHisHisArgHisAaPGLYArgAaPSerArgArg 518
QY 936 ---CTCCGACGCGCGCGCGCGAGCTCCGCTGAGTTGGGC----- 974
DB 519 ArgValProAaPProArgArgAlaGlyProProGlyArgLeuProProGlyArgProGlnArg 538
QY 975 -----CCGAGCGCTGGGCTC-----AGCGCGGCGCTCGG 1004
DB 539 ProAaPProValProAlaGlyGlyArgProGlnGlnSerAlaGlyArgArgProLeuArg 558
QY 1005 CTACCACTGTCAGTTTGGGCGCTGCGGAGCGGAGGCGCAGCGGAGAGCTGCCCGC 1064
DB 559 LeuProAlaSerGlyArgAlaAlaCys-----ValGlnGlnGlyGlyArgArg 574
QY 1065 GGGCGCGCACTGCACACGCT----- 1085
DB 575 CysProProGlyArgGlnAlaGlyGlnHisLeuArgArgAaPGLYArgArgSerArgGly 594
QY 1086 -----GCAGGGCTGTGGCGCGCGGCGCATCTGAGGACGTGACCGCGCTTCCG 1139
DB 595 CysArgAlaAlaAaPGLYAaPGLYLeuArgGlnProValAaP---GlnArgHisGlnPro 613
QY 1140 CTC-----GGCGCGGAGAGCGCGCTGCCG 1166
DB 614 AlaGlnGlyGlnValAlaAlaAlaProAaPHisProGlyGlnGlyProGlyProAlaArg 633
QY 1167 CAAAGCTGTCAGCGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1217
DB 634 ProValAlaHisLeuArgGlnProAlaGlnHisProGlnLeuAlaAlaProGlyValAla 653
QY 1218 GCGCCCGACCGCGGGGAGCGC---CAGCGG 1244
DB 654 GlnProArgProGlySerArgAaPGLNPro 663

RESULT 10
US-09-252-991A-31760
; Sequence 31760, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31760
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31760

Alignment Scores:
Pred. No.: 1,88e-10 Length: 1706
Score: 269.00 Matches: 153
Percent Similarity: 32.2% Conservative: 19
Best Local Similarity: 28.7% Mismatches: 198
Query Match: 11.3% Indels: 164
Gaps: 30

US-10-699-035A-5 (1-1254) x US-09-252-991A-31760 (1-1706)

QY 1246 CACGGCTGGCGGTCCTCCCGGGGTCGGGGGCGACCGGGGCGTGGCGCGGGCGCGGCG 1187
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 31 HisgLy-----ProGlyAspArgGlyPro-----AlaProAlaGlu 43
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1186 CGTCGGGGGTGACGGCTTTGGGGGACAGCGCGCTCTCGGGGCGCGAGCGGAAAGCGCGCG 1127
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 44 GluProAlaAlaLeuArgHisArgThrGlyAspArgAlaAlaProAlaArgGluArg 63
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1126 TCACGGTCACCA----- 1115
Db :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 AlaGlySerProArgGluProAlaGlyValAlaAlaLeuLeuHisArgGlyGlnArgPro 83
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1114 -----GGTAGCGGTGCGCGCGCGCGCGCGCGCTGCGAGCGTGTGCGAGTTCGCGCG 1064
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 84 ValProGlyGlyArgArgAlaAlaGlyProAlaGluProGlyAlaAlaGlyGlyPro 103
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1063 ----- 1049
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 104 AlaArgHisArgGlnGluProAspHisArgGlnHisGlyArgArgArgHisArgProPro 123
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1048 GCTGGCGCTCC-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1010
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 AlaGlnProAlaAspArgLeuProGluAlaArgArgProGlySerGlyAlaGlnAlaPro 143
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1009 GGTAGCCGAGCGCG-----CGGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 962
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 144 GlyGlyArgGlyProGlyProAlaHisArgHisAlaGlnArgGlnArgProArg 163
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 GGAAGC-----TGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 914
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 164 ThrGlyAspProGlnHisProArgAlaAlaArgSerAlaValArgArgSerArgArg 183
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 913 CTTCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 857
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 ProGlyLeuGlyAlaProAlaArgAlaProGlyGlyAlaTyrArgGlyAlaGlyArgArg 203
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 856 ACTCAGGAGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 800
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 204 ThrGlyProLeuProProGlyProAlaPro--ArgGlyAlaArgGlyHisArgAlaGluLeu 222
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 799 ---AGTCGCGCGCTTCCTGGCAGCTGTGGCGCTTCGCGCGCGCGCGCGCGCGCGCGCG 743
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 223 ProAlaProAlaArgArgThrAspArgAlaGlyGluArgLeuAlaAlaArgLeuPro 242
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 742 GCACAGAGCTCAGCAGCATGTAGTCCGAGCTGCGCGCGTGCAGGAGGTGCAGGCGCGCGCG 683
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 243 GlyProAlaProAlaProGlyGlyAlaGlyHisArgGlnProGlyAla-----ProGly 260
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

QY 682 GGAAGCGCTGAGCTGATCTCCG-----TGGCARGAGCT 647
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 261 GlyIysLeuArgAlaAlaGlyProProLeuAlaThrGlyProLeuArgArgGlnProAla 280
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 646 GCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 587
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 261 GlyAlaThrAlaArgLeuArgHisArgPro-----GlyHis 292
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 586 CCA-----CGTCCCAAAAGTGCAGGCTCTTCGCGAGGGGCTGAGCGAGCGCTG 536
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 293 ProAlaArgLeuArgGlnProAlaArgLeuArgArgSerGlyGluArgProAlaGlu 312
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 535 ACA-----GCTCCAGGAAGTTCCTCGCGCGCGCGTCTGACATGAACGCGGAGCG 485
Db :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 313 SerProArgArgThrProGlyGluLeuArgGlyArg-----Arg 325
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 484 CCAGGCTCTTAGCTCTGCTGATGGGGGCGCGACAGGCTCGTG-----AGC 437
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 326 ProAspAla---LeuProArgLeuAlaGlyIysProGlyArgGlyProAlaGluPheAla 344
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 436 CGCGCATCTGTACCCACACACAGCATTTGGGACCGCTGGCGCGGCGACCTGATGCTTCAG 377
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 345 ArgGlyAlaAlaProAlaPro-----GlyProLeuAlaAlaValProArgArg 361
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 376 -----CAACAGCTCTTCTTGGCATGACCGCGCGCGCG----- 341
Db :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 362 HisAlaArgAlaAlaGluProThrArgArgProAlaGlyGlyArgAlaAspArgThr 381
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 340 -----CAGTGTGGGTGACACCATGC-----GCTGGGCGAGAGAC 305
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 382 AlaThrAlaProAlaArgProArgThrAlaCysAlaAlaAlaValAlaGlyProArgHis 401
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 GCACCGCATCTCTGGGCGAG-----CTTCACCGAGCTGTGCTGGCGCGAGGAGAACT 254
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 402 AlaArgArgGluGlnProArgThrLeuProGlyProArgProProGlyAlaArgGlyThr 421
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 253 CGGTGTATGGCGCGAGTGC-----CCAGTGCACACAGACTGACGCGAGGGCGCGCGCGCA 197
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 422 HisLeuAlaGlySerAlaGlnProAlaAlaPro-----AlaAlaAlaArgProPro 438
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 196 GGGGCGAGT-----GAGCGACCA----- 179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 439 AlaAlaLeuProAlaArgAlaArgArgGlyArgArgArgGlnProProGlyArgProPro 458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 178 -----GCTGCCCAAACT----- 164
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 459 ArgArgLeuArgProGlyAlaProValAlaAlaAlaProAlaThrGlyGlyAlaAla 478
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 163 CCCGAACCCGGGAGAACTCGTAGAGAGAGCGTGGCTGCTGCCAGGAACTCA 104
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 479 ThrGluProGlyAlaGlyAspSerArgHisArg-----CysArgSerGlyThr--- 494
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 103 GGTCCCTCGGGGGGCTGATCTGTGGACCGCGCTCCGCGCGCTCCGCGCGAGCGCA 44
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 495 -----AspArgProGlyArgProArgGlnProAlaAla 505
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 43 GCCCGAAGCTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 506 AlaArgAspProGlyProGlySerThrProGlyProSerAla 519
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-09-252-991A-30843
; Sequence 30843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
```



```

QY 1126 -----TCAGGATACCAAGTAGAGCGGTGCCCGGCGCAAGCCCTTCAGCGGTGTGCAG 1074
DB 114 yValValGlyLeuGlyProGlyGluArg-----ProGlyProAlaPheLeuSerAr 131
QY 1073 TTGCGGGC----- 1067
DB 131 gGlyGlyLeuGlnAraArgArGlyAraSerArGlyLeuSerGlnAlaAspArgLyHis 151
QY 1066 -CCGCGGAGCACTCCACCCGCTGCGCTCCGCGCCGCGAGCGGCCGAAGTCAGCGTGG 1008
DB 151 sProProArGArGProAlaAlaHisProSerArGPro-----ArgArGThSerGlnG 169
QY 1007 TAGCCGAGCGCGCGCTGAGCCGAGCGCTGGGGCCCACTCAGCGGAGCGCTGCGCGCG 948
DB 169 y-----AlaGlnArGAlaValProGlyGlyValProArGAlaArgProGly----- 184
QY 947 CGGGGCTGGAGATAGACGATGGCTGGCCCGGCTCTGCGCGCGCTGCGCGCACCCGCG 888
DB 185 -----GlyArGLeuAlaLeuAlaGlyLeuSerTTrpProGlyHisGlySerArGProAl 202
QY 887 AGGATCTGGG-----GCCTCAGAGAGCGCACGTTGGACTCAGAGCACTAGCGCGCAG 837
DB 202 aGlyThrGlyAraPAlaAspAlaArgHisAlaGlnArGHisGlnArGAlaGlnAlaProG 222
QY 836 TCGTAGTCCGCTGCGCGGTGAGGC-----CGGCGCGAGATCCAGTCC 795
DB 222 yArGAlaAlaAlaThrGlyArGAspSerValArGlyGlyGlnArGHisAlaArgAspR 242
QY 794 GTGGGCTTCCCTGGCAGCTGCTGGC----- 770
DB 242 aArgSerGlyLeuGlnProAlaGlyGlyGlnProGlyLeuGlnSerHisAlaArgLeuSe 262
QY 769 -----GTCTTGACGCCCGCGGCTGGCGGTGG-----GC 741
DB 262 rProArGlyArGlyArGAlaGlnArGlnProAlaGlyGlyTyrProArGlyProAl 282
QY 740 ACCAGCTCCAGCA-----CATATGAGCCGAGCTGCGGTGTCAGAGGGGTGGCCAG 690
DB 282 aProLeuProAlaAspProArGArGAlaAlaProAlaArGAlaValAlaGly-----Ar 300
QY 689 GCCAGCGCGA-----AGCCGCTGG---ACGTGATCTCCGTGGCATGAGCTGCTCGGC 639
DB 300 gThrGlyGlyTyrAlaGlnGlyLeuTTrpArgSerLeuSerProMetAlaAlaAlaGlnProAl 320
QY 638 CCGATCGAGAGATGAGCCCTCAGCTTGGACAATGATGAGAGTCA----- 587
DB 320 aAlaArGArGlnTTrpArGlyAspPro-----HisArGArGPh 333
QY 586 -----CCAGCTCCACAAAGTGCAGGTTGTTCTCGGAGGGGCT 549
DB 333 eleuHisArGSerGlnCyseProProArGArGArGlyAlaProAlaSerProAlaAlaLe 353
QY 548 GAGGACGCGCTGACAGCTCCAGAAATTGCTCGGCGCGTCTGACAATGAACAGGTG 489
DB 353 uArg-----ArgThrAspArgProGly----- 360
QY 488 AGGCCCAAGT-----CCTTGAGCTCTGATGGGGGGGCCCAAGGGTCCGTGAGCGCG 435
DB 361 -GlnProProValAlaProArGAlaAlaAlaArGlyGlyProAlaGly----- 376
QY 434 CCATCTGCACCCAGACAGCACTTTGGGCAACCCCTGGCGGAGCACTGATGTTTCAGCA 375
DB 377 -----AlaProGlyGlyGlyArGAlaGlyProAlaAla----- 387
QY 374 AACAGCTGTTCTTGCGCATAGACAGCGCCAGGCGCAGTGT----- 335
DB 388 -----AspArgProGlyProLeuGlnAlaAlaGlnArGlnProArGPr 402
QY 334 -----GGGTGTACCCATGTGGCTGGGAGAGACAGCGCATTCCT----- 293
DB 402 oArGlyArGArGProValAla-----GlnGlyAspGlyProAlaProAlaAlaHisArGAl 421

```

```

QY 292 -GGGAGGCTCACCCGAGCTGTGCGCCGAAGGGAACTCGGTGTATGCGGACTGCC 234
DB 421 aGlySerGlnHisProGlyAlaSerGlyArGArGValArg----- 435
QY 223 AGGTGCACCAAGCTGG-----CAGCGAGGGCGCGGTGTCGAGGGGAGTGAAGCC 183
DB 426 -HisSerProArGTrpArGHisArGHisArGHisArGHisArGHisArGHisArGHis 453
QY 182 ACCAGCTGCCCCCAAACTCC-----GAACCCGGAGAACTCGTAGAGAGACGCTG 129
DB 453 rProAlaGlyProValAlaProAlaGlyGlnArGAlaArGProArGValaAspProArGAr 473
QY 128 GCTGAGC-----TGTCAGCAGAAACATCAGGT----- 101
DB 473 gLeuAlaGlyAspGlnProValPheArGAlaGlyThrGlyAsnLeuArGAlaAspGlyPr 493
QY 100 -----CCCTCGGGGG 90
DB 493 oGlyGlyProGlyAraAlaThrCyseGlnAlaThrArGArGlnTyrLeuProValLeuPr 513
QY 89 GCTGATGCTGTGACCGCGCTCCGCGCGCTCCGCGCCAGCG---CCAGCCGAAGCTC 33
DB 513 oArGlnProAlaGlyGlnHisProArGThrProAlaProGlyAsnProValAlaGlnG 533
QY 32 AGGG----- 15
DB 533 yGlyArGlyGlyProAlaGlyArGAlaLeuProAlaGlnAlaValProGlyGlnArGPr 553
QY 14 G-----TCCAGGGAGCA 2
DB 553 oProGlyLeuGlyArGlyAla 560

```

Search completed: February 13, 2006, 13:44:00
 Job time : 72.2111 secs

THIS PAGE BLANK (USPTO)